Crosstalk of synthetic cassettes in defined chromosomal sites

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina

zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Shawal Spencer
aus Jagadhri, India

1. Referent: Prof. Dr. Reinhard Köster
2. Referent: Prof. Dr. Michael Rehli

eingereicht am: 26.05.2014
mündliche Prüfung (Disputation) am: 23.07.2014

Druckjahr 2014

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Kruse, N., Spencer, S., Wirth, D. . (in press (2014)) In Hauser, H., Wagner, R. (ed.), *Animal Cell Biotechnology in Biologics Production*. DeGruyter, Berlin

Tagungsbeiträge

Spencer, S.: Crosstalk of synthetic cassettes in defined chromosomal sites- An Epigenetics Perspective (Oral presentation). COST conference: Epigenetic: From Bench to Bedside Athens, Greece (2014).

Spencer, S.: Crosstalk of synthetic cassettes in defined chromosomal sites- An Epigenetics Perspective (Oral presentation). Progress Seminar, Helmholtz Centre for Infection Research, Braunschweig (2014).

Spencer, S, Zha L,:Towards An HCV mouse model: Overcoming Epigenetic Roadblock(Joint oral presentation). 4th Public retreat of the HZI Graduate School, Hahnenklee (2013).

Spencer S.: Crosstalk between synthetic cassettes in a defined chromosomal site(Oral presentation). ESACT Conference, Lille, France (2013).

Spencer, S, Natascha Kruse, Hansjörg Hauser and Dagmar Wirth: Crosstalk between synthetic cassettes in a defined chromosomal sitePoster presentation). 3rd Public retreat of the HZI Graduate School, Bad Bevensen (2012).

Spencer,S.: Crosstalk of synthetic cassettes in defined chromosomal sites(Oral presentation). Progress Seminar, Helmholtz Centre for Infection Research, Braunschweig (2012).

Acknowledgement

I would like to thank Dr. Dagmar Wirth for accepting me as a PhD student in her department and for her honest and generous support during all these years. She kept patience and calmness all the time and under her leadership I had the chance to learn many new things. I would like to thank Prof. Dr. Reinhard Koester for accepting to supervise and review my thesis. I am very grateful to Prof. Dr. Michael Rehli for his enthusiasm, his continuous support and advice, for accommodating me briefly in his lab and his lab members who made my short stay in Regensburg a thoroughly enjoyable one. I am also grateful to Dr. Hansjörg Hauser. He was always willing to listen to my problems and was very supportive throughout. I would like to thank my Grad school for their continuous support and enjoyable retreats.

I would like to thank everybody in MSYS for the wonderful atmosphere, their patience and support. In addition, I would like to take the opportunity to thank our closely connected co-workinggroup – the RDIF department and EXIM colleagues. We not only shared room and working material, I also appreciated our combined seminars, brainstorming where I got a chance to learn and widen the span of my knowledge. I also would like to thank the whole sequencing department and also a special thanks to Maria and Lothar for their technical support in difficult FACs sorts and analysis.

Special thanks to Dr. Natascha Kruse for her precious organization skills and help in the various difficult situations (on numerous occasions). Thanks to Sara for her technical and moral supportespecially when it came to ordering mice.

I would also like to thank all former lab members whose support made the journey a memorable one: Dr. Daniel Maeda, Dr. Kristina Nehlsen, Agustina Gugliotta, Petra Voeller, and Sabrina Herman. I would also like to thank all other friends not personally mentioned yet who helped to make my stay in Braunschweig fun and memorable one.

Last but not least, I would like to thank my parents and my younger sister. Their support, motivation and love helped me first to start this journey and then to make it all the way until the end. Without their help I would never have had the chance to start and finish my study, to work in and explore Germany for over three years and to perform the present thesis. I really appreciate their never-ending love, help, support and patience.

Table of Contents

ABSTRACT

Heterogeneity in transgene expression is frequently observed upon genetic modification of cells in basic research and biotechnology. The variable transgene expression is considered to be a result of the crosstalk of the incoming promoter cassette with cis-acting elements associated with the chromosomal site of transgene integration (position effect). Targeted integration of the transgene into the open chromatin reduces variability due to chromosomal position effects and also favors the more predictable transgene expression. The objective of this study was to have a mechanistic understanding of the nature of interaction that occurs between the transgenic/synthetic cassettes when integrated into defined chromosomal sites. To this end CMV driven transgene expression were investigated in more than 100 independent cell clones in two different cell lines (CHO and HEK293T). Not only was the transgene expression highly variable among clones but also large levels of heterogeneity in expression existed within clones with metastable phenotype. This was correlated with differential and dynamic chromatin conformation related to differential histone modifications. In addition, a CMV based Tetracycline inducible synthetic promoter (BiTet) was evaluated in the well-characterized Rosa26 locus.The epigenetic status of the promoter cassette was evaluated in mouse ES cells and transgenic mice to investigate the mechanisms mediating the interaction between the transgene and the chromosomal loci that results in variation of transgene expression. Contrary to the expectation, even upon targeting the ubiquitous Rosa26 locus, the expression of the synthetic cassette driven by tetracycline inducible synthetic promoters (BiTet) was highly heterogeneous. However in this analysis the endogenous Rosa26 locus largely remained methylation free. While DNA methylation was the major player in the silencing of the Tetracycline based promoter systems in the Rosa26 site, the heterogeneity associated with the hCMV driven constructs in random CHO and HEK293T clones was entirely associated with distinct histone modifications causing variable transgene expression and transgene silencing. While the heterogeneous expression was found to be associated with different chromatin states conferred by various epigenetic markings, the stability and extent of the variation in transgene expression may largely depend on the nature of crosstalk between the chromosomal integration site and the synthetic construct.

ZUSAMMENFASSUNG

Genetisch manipulierte Zellen werden oft in der Biotechnologie und der Grundlagenforschung verwendet. In diesen genetisch manipulierten Zellen kommt es oft zu unerwünschten heterogenen Transgen-Expressionen, die oft ein Nebeneffekt von sogenannten Positionseffekten sind. Diese Positionseffekte entstehen durch die Interaktion der Promoter-Kassetten mit den im Genom codierten Cis agierenden Elementen, was zu unterschiedlichen Transgenexpressionen führt. Die gezielte Integration eines Transgens in einen offenen Chromatin-Abschnitt reduziert den sogenannten Positionseffekt und führt zu einer vorhersagbareren Expression.Das Ziel dieser Arbeit ist die Untersuchung von transgenen/synthetischen Kassetten an verschiedenen chromosomalen Integrations-Orten um die sogenannten Positionseffekte und den chromosomalen „crosstalk“ genauer zu charakterisieren. Von zwei verschiedenen Zelllinien (CHO und Hek293T) wurden 100 Klone analysiert. Diese Klone wurden auf die Expression der integrierten Transgen-Kassette überprüft. Hierbei wurde das Transgen von einem CMV Promoter exprimiert.

Zum Einen zeigte die Analyse der Zellen, dass die einzelnen Klone mit der gleichen Transgenkassette an verschiedenen Integrationsorten unterschiedliche Expressionsmuster aufwiesen (interklonale Expressionsunterschiede). Interressanterweise änderten sich zum Anderen diese Expressionsmuster der analysierten Klone nach mehrmaligen passagieren (Intraklonale Expressionsunterschiede). Das heisst, dass Zellen mit der gleichen Transgenkassette an dem gleichen Integrationsort ihr Expressionsmuster verändern.

Diese intraklonalen Unterschiede wurden als metastabil bezeichnet. Die unterschiedlich exprimierenden Zellen in den metastabilen Klonen wurden genauer charakterisiert. Hierbei korrelierten die verschiedenen Expressionsmuster mit den unterschiedlichen Histon Modifikationen und somit mit den chromatin Konformationen. Zusätzlich wurde ein CMV basierter synthetischer Promoter, Tetracycline abhängiger Promoter (BiTet), in einem gut charakterisierten Lokus, dem R26 Lokus, analysiert. Obwohl der Promoter (BiTet) in diesem ubiquitär aktiven Lokus integriert wurde, zeigte die Expressionsanalyse, sowohl in Maus-Stammzellen als auch in transgenen Mäusen, überraschenderweise eine heterogene Expression. Anhand der epigenetischen Analysen konnte gezeigt werden, dass der Bitet

Promoter methyliert wird, wobei der endogene R26 Promoter hauptsächlich frei von Methylierungen bleibt.Während das Silencing der Bitet Kassette in dem R26 Lokus und somit die heterogene Transgenexpression hauptsächlich auf die Methylierung der DNA zurückzuführen ist, ist das Silencing der analysierten CHO und HEK293T Klone, die zufällig im Genom integrierten, ein Resultat der Histon Modifikationen.Nichts desto trotz, während die Heterogenität der Transgenexpression von verschiedenen Chromatin-Zuständen, die über bestimmte epigenetische Markierungnen etabliert werden, abhängt, ist die Stabilität und die Stärke der heterogenen Expression abhängig von der Interaktion des synthetischen integrierten Konstrukts und des chromosomalen Integrations-Ortes.

1Introduction

1.1 Epigenetic role play in gene expression

It was long held that the enormous wealth of information stored in the genome sequence of an individual's to a large extent determines the phenotypic outcomes. The human genome project was thought to solve many riddles, however, that data suggest that only about 30,000 genes are involved in the functional existence of the most complex organisms on the planet. In spite of the abundance of the information it is still quite vaguely known how this information is managed, released and maintained. Studies have shown that inspite of the same complement of genetic material incase of twins, the phenotypic variations have been large. Such discordance between the genotype and phenotype has been associated with the epigenetic modulations of the underlying genotype.

These modulations are considered to be responsiblefor the alteration in the function of genes without a change in the genome sequence. The field of research dealing with such kind of mechanisms is known as "epigenetics", and involves chemical modifications like methylationoccurring on the cytosine residues of the DNA and also modifications of histone residues such as acetylation, phosphorylation and methylation (Figure 1). The term "epi" means "above/over", suggesting the location of the modifications with respect to the DNA sequence. The epigenetic modification forms a kind of shell covering as well as interacting with the core information in the genetic sequence in the body. This interaction results in additional information that governs the differential expression of genes.These epigenetic mechanismsplay an important role during the development stages in an organism where about 200 different cell types sharing an identical genotypearise from a single cell. Its importance in the differentiation process is also evident from the fact that different progenitor cells, again sharing same genotype, form a variety of cell types with diverse, yet stable, profiles of gene expression and distinct cellular functions.

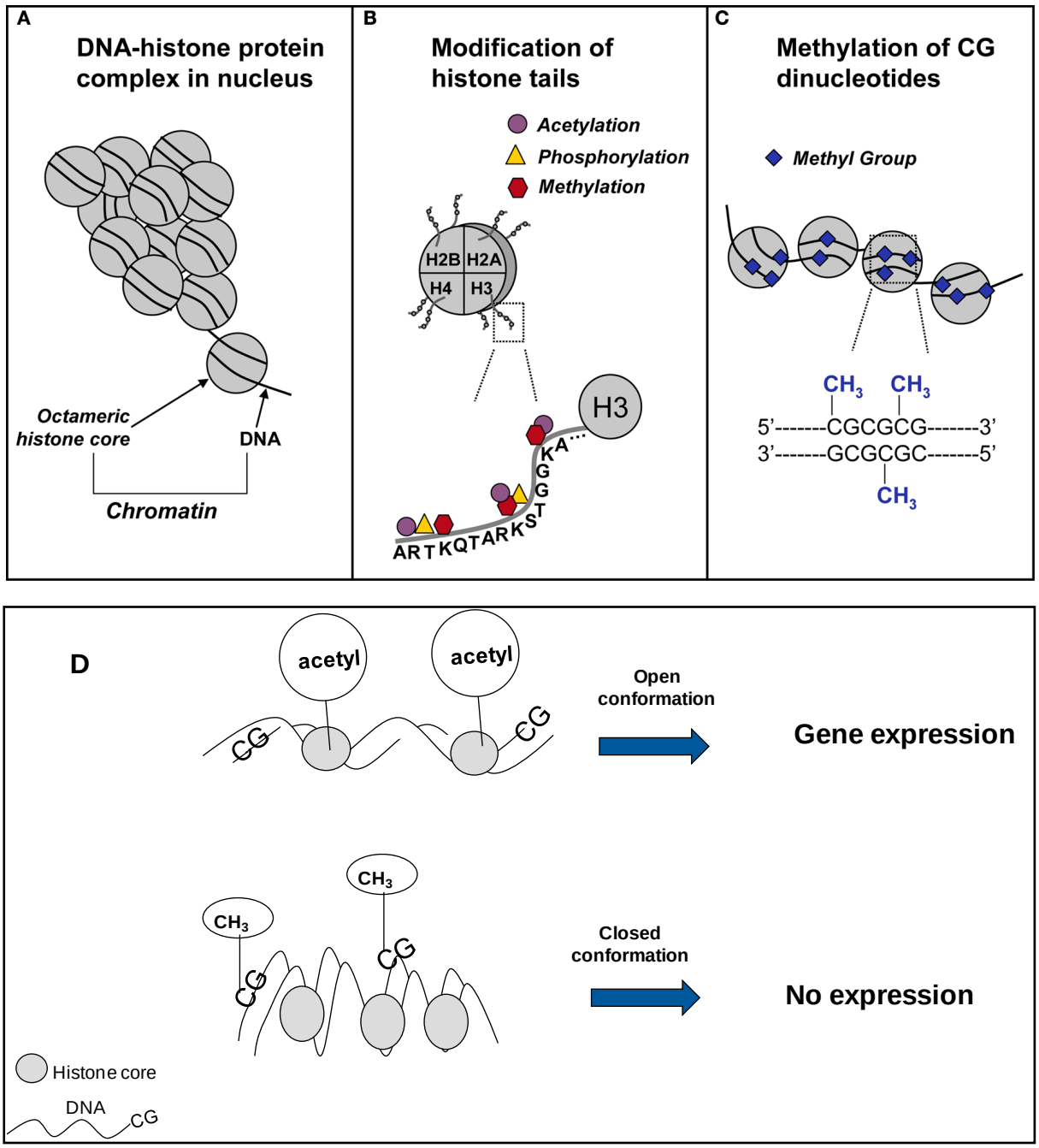


Figure 1: Epigenetic regulatory mechanisms. A) The organization of DNA with histone. B) Modifications of histones H3 tail residues such as lysine (K) and arginine(R). C) In addition,cytosines of the CpG motif undergo methylation on the DNA level (figure adapted from Penner et al(1)). D) Epigenetic modifications such as acetylated histones and demethylated cytosine favor open conformation thereby favoring transcription. Deacetylated histones and methylated cytosine favor closed conformation thus repressing transcription of genes.

The differential epigenetic markings on DNA and histones alter the conformation state of chromatinEpigenetic marks like acetylation of histones, trimethylation of histone H3 lysine 4, unmethylated cytosine etc make chromatin more accessible for transcription factors to bind and thereby supporting the gene expression.On the other hand epigenetic marks such as methylation of DNA,deacetylation of histones, trimethylation of histone H3 lysine 9 and histone 3 lysine27 are some of the epigenetic marks present at the condensed chromatin. The condensation of chromatin thus preventing the binding of transcription factors leading to silencing of gene expression(2-6).

The following chapters will focus on DNA methylation and some modifications of histones in more detail since they represent well characterized mechanisms and form an integral part of epigenetic regulation.

1.2 DNA methylation

One of the most well characterized and known epigenetic modification is that of DNA methylation that involves covalent addition of a methyl group to the 5' carbon atom of cytosine (C)(7). In mammals, this modification mainly occurs to cytosines that are succeeded by guanine (G), thus forming the CpG dinucleotides (CpGs).These CpG dinucleotides are not uniformly distributed. Rather, they occur in clusters, thereby forming a kind of islands on the otherwise CpG devoid genome (so-called CpG islands, CGIs) (8-11). One plausible explanation for non-uniform distribution of CpGs might be that methylated cytosines(5mC) undergo hydrolytic deamination yielding thymine (T). Thymine is a naturally occurring genomic base and is not recognized by the repair system of the cells. Thus, deamination of 5mC results in a C to T transition. This might have accounted for the asymmetric distribution of the CpGs during the course of evolution.

The classification of a region into a CpG island is based on the following criteria: 1) an overall GC content of 50 % or higher 2) The observed versus expected ratio of CpG frequency of 0.6 in a region of 200 base pairs minimum(10). Interestingly, CpG dinucleotides are distributed unequally over the genome.In most cases methylation of the CpGs is associated with gene silencing and is considered to play an essential role in embryonic development (11), cell differentiation and dedifferentiation (12), genomic imprinting(13,14), X-inactivation in mammals(15)and silencing of potentially harmful DNA elements like transposons or endogenous retroviruses(16).

The addition of the methyl group from the methyl group donor (mostly S-adenosyl methionine, also abbreviated as SAM) to the 5’carbon of cytosine is catalyzed by enzymes called DNA methyltransferases (Dnmt) (Figure2B).In eukaryotes, three different families of DNA methyltransferases have been identified, namely Dnmt1, Dnmt2 and Dnmt3 with the two subclasses 3a and 3b(17). Although the mechanism of action is similar, they differ in their biological roles. Dnmt1 is responsible for maintaining the methylation pattern during the replication in the S phase of the cell cycle. It copies the methylation marks onto the daughter strand from the hemimethylated parent strand (Figure 2C). It exhibits a 5-10 fold higher affinity for hemi-methylated DNA (18,19). The functional role of Dnmt2 has not been clearly elucidated, however, it is proposed that Dnmt2 might be playing a role in RNA methylation (20,21).Dnmt3a and Dnmt3b are responsible for de novo DNA

methylation(Figure 2C). They become active during early stages of developmentand are essential in the establishment of new methylation patterns and following correct development (22,23). Dnmt3L does not have a catalytic function of its own but studies have shown that it also plays a role in catalytic activity of de novo methyltransferases(24)

With the growing importance of DNA methylation mediated gene silencing and its role in cancer and other diseases (17,25), the search for the inhibitors of DNA methyltransferases lead to the discovery of chemicals like Azacytidine and Decitabine. These were then followed by the identification of various different chemical drugs. However, these two remained the most widely used Dnmt inhibitors. . They represent nucleoside analogues and upon phosphorylation in the cells, they are incorporated into the DNA or RNA(Figure 2D)(26). While the incorporation of Azacytidine has been shown to be preferential for RNA, Decitabine prefers the DNA(27). Mechanistically, Azacytidine (Aza) and Decitabine (Deci) get incorporated into the DNA after being metabolized to 5-aza-2′-deoxycytidine-triphosphate which substitutes for cytosine. Azacytosine-guanine dinucleotides are recognized by the DNA methyltransferases as a natural substrate. Dnmt enzymes will initiate the methylation reaction by a nucleophilic attack. This results in the establishment of a covalent bond between the carbon-6 atom of the cytosine ring and the enzyme. The bond is normally resolved by beta-elimination through the carbon-5 atom, but the reaction is blocked with azacytosine, where carbon-5 is substituted by nitrogen. Thus, the enzyme remains covalently bound to DNA and its DNA methyltransferase function is blocked. In addition, the covalent protein adduction also compromises the functionality of DNA and triggers DNA damage signaling, resulting in the degradation of trapped DNA methyltransferases. As a consequence, methylation marks become lost during DNA replication. With repeated divisions DNMTs are blocked by being bound to these drugs. As a consequence of this, significant demethylation can be observed.

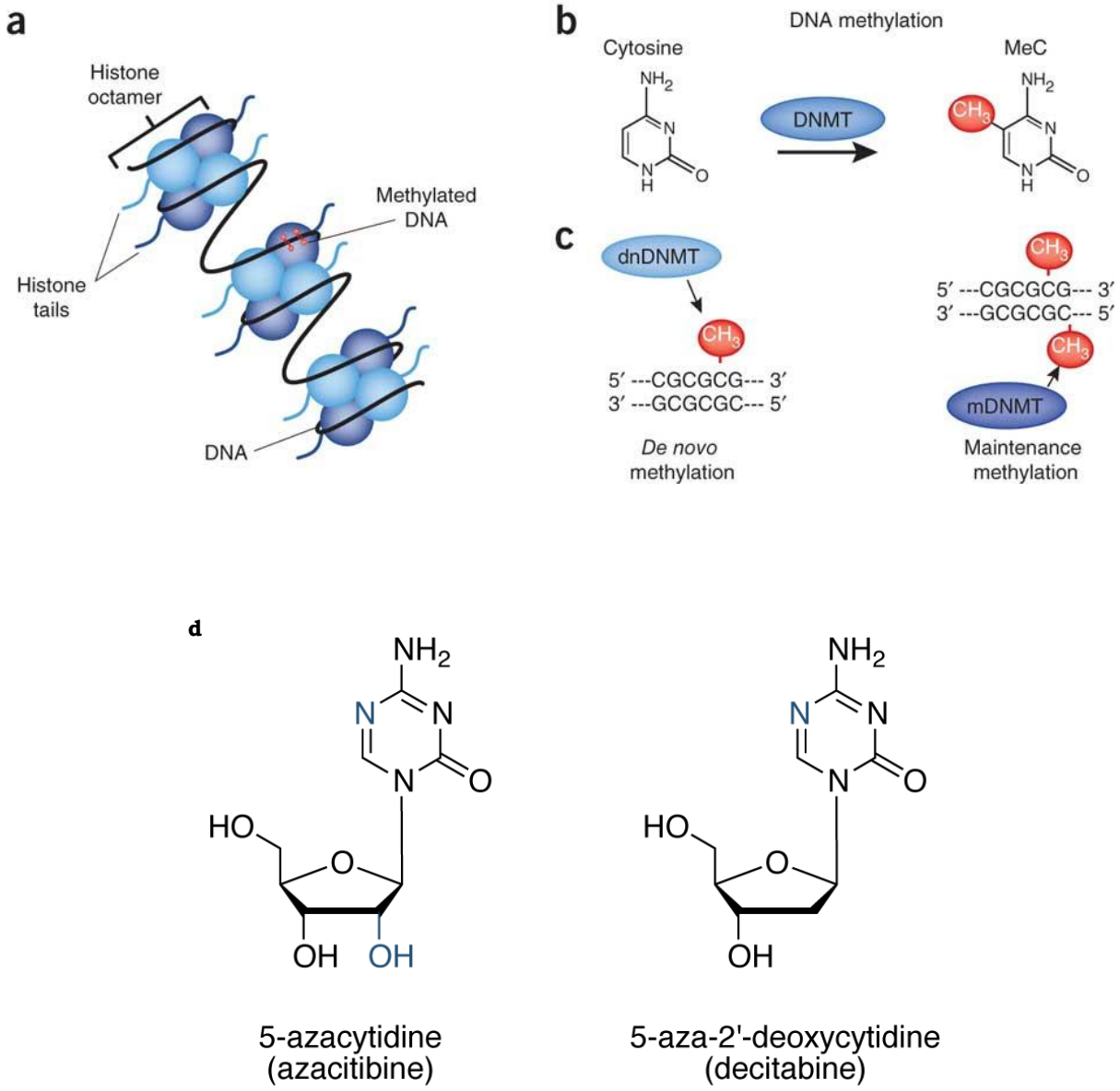


Figure 2: DNA methylation reaction catalyzed by Dnmts. A) Presence of DNA methylation on the chromatin. B&C)the DNA methylation reaction catalysed by DNA methyltransferase (denovo(dnDNMT)) and maintaince(mDNMTs) in presence of S-adenosyl methionine (SAM) (Figure adapted from Day et al(28)). D) Structures of Azacytidine and Decitabine (Dnmt inhibitors)

1.2.1 Role of DNA methylation in healthy cells, in transcriptional regulation and in cancerous cells

DNA methylation is an ancient process found in all domains of life. In eukaryotes, which include diverse organisms such as plants as well as humans, DNA methylation is found exclusively at cytosine residues (29-31). Several studies have indicated the essential role of

DNA methylation in proper development of a multicellular organism(32). DNA methylation has been proposed to be an important role player in mammalian development in establishing the correct pattern of gene expression. Whether it is X chromosome silencing or transposon silencing, the role of DNA methylation has been shown to be imperative(16,18). The observation that mice lacking Dnmts die very earlyduring embryogenesis suggests that this modification has important roles and is essential for mammalian embryonic development(33).Embryos show reduced DNA methylation levels but the specific reasons for death during development remain unclear. Defects in repression of the inactivated X chromosome in female cells and in the establishment and maintenance of allele-specific expression of imprinted genes have been observed(34). Thus, lethality might result from aberrant gene dosage.

The role of DNA methylation in a normal cell can also be evaluated on the basis of the exact location of the methylation. The promoter methylation of oncogenes has been associated with the inactivation of oncogenes whereas aberrant methylation can lead to uncontrolled proliferation resulting in cancerous growth(35). Also the promoter methylation plays a role to suppress the activity of lineage specific genes allowing for specific expression of genes in a lineage specific manner.

High levels of DNA methylation are also associated with Alu repeats elements that are kept in a silenced state. In *Dnmt*-knockout mice, global demethylation as a consequence of Dnmt1 absence likely triggers mutations through the activation of cryptic transposons, which might contribute to early lethality (36,37). Normally these transposons are methylated and thereby repressed. Consequently, DNA methylation masks the effects of transposon insertion by mechanisms that do not directly depend on regulation of transcription or transposition.

However, the evolution of DNA methylation has been found to be more ancient in case of the gene bodies and intragenic regions. Also in most of the invertebrates, DNA methylation predominantly occurs in the gene body regions (7,16,30,31). These studies have also suggested that the gene body methylation might have evolved before the promoter methylation. The gene body methylation has been hypothesized to reduce transcriptional

noise (38,39). These transcriptional noises might be due to the phenomenon of transcription occurring in patches rather than continuously, a process called transcriptional bursts (16,40,41). Another reason for the noise can be transcription from non-canonical promoters, potentially due to overabundance of RNA polymerase II in the cellular environment(16) which are thereby repressed by DNA methylation.

The study by Huh et al found a negative correlation between the gene body methylation and transcriptional noise supporting the hypothesis that the gene body methylation evolved to suppress the transcriptional noise (16,31) thereby reducing the production of transcripts from non-canonical promoters and reducing the waste consumption of energy. The same group showed that differential gene body methylation in prefrontal cortex and blood resulted in significantly different transcriptional noise with less transcriptional noise in brain (more methylated) than blood(less methylated) (16).

Cancerous cells have been characterized by aberrant methylation profiles with general hypomethylation in the intergenic regions and at oncogenes. However, they provide a hypermethylated state at the tumour suppressor genes, thus silencing them and creating a perfect environment for unchecked growth due to expression of Hypomethylated oncogenes.

1.3 Histone mediated gene modulation

Histones form the main protein component of the chromatin with DNA wrapped around them. The basic unit of chromatin, the nucleosome, is formed by a histone protein octamer. This octamer is formed of two molecules of each of four core histones H2A, H2B, H3 and H4 with 147bp of DNA wrapped around(42-44). Various modifications of histones have been found on different amino acid residues. These include phosphorylation of threonine, serine and tyrosine(45). Also arginine can undergo mono, di, or tri methylation. However, most commonly studied modifications include the acetylation and methylation of lysine residues. All these modifications can have a different role in gene expression depending on the kind of modification and on the particular residue being modified. In general, acetylation of histones have been associated with the active gene transcription where as methylation can

have both activating (H3K4me3) as well as repressive (H3K9me3 and H3K27me3) effect on the chromatin(45).

1.3.1 Crosstalk between histone modifications and transcriptional regulation

The presumption that a specific histone modification has a specific effect on transcriptional regulation (activation or repression)might hold true for modifications like acetylation which is activating and SUMOylation which is generally repressing (46). However, in most other cases the scenario is more complex with evidences of a complex crosstalk existing between different histone modifications. For example in spite of being mutually exclusive and having completely opposite roles, both H3K4 methylation and H3K9 methylation were shown to have a crosstalk where the increase in levels of H3K9 methylation leads to decrease in H3K4 methylation and vise-versa(47). Another interesting case is that of H4K20 methylation where the extent of methylation determines the overall acetylation levels on H4. In this case the monomethylation of H4K20 leads to hyperacetylation and hence gene activation whereas the trimethylation of the H4K20 leads to gene repression (44,48,49).

Thus the histone mediated gene regulation presents an example of complex network of regulatory modifications that may function in a highly ordered manner depending upon the crosstalk that occurs between them. Most common histone markings and their effect on gene expression is summarized in Table 1.

Type of histone modification	Abbreviation	Function
Histone3lysine4trimethylation	H3K4me3	Gene activation
Histone3lysine4 acetylation	H3K4ac	Gene activation
Histone3lysine9trimethylation	H3K9me3	Gene repression
Histone3lysine27trimethylation	H3K27me3	Gene repression
Histone3lysine27acetylation	H3K27ac	Gene activation
Histone3lysine20trimethylation	H3K20me3	Gene activation
Histone3lysine36trimethylation	H3K36me3	Gene activation
Histone3lysine79 dimethylation	H3K79me2	Gene activation
Histone3lysine79trimethylation	H3K37me3	Gene activation

Table1: Summary of the most common histone markings and their respective association with the gene expression.

1.3.2 Histone code hypothesis

The histone code hypothesis was suggested in 2000 by Stahl and Allis, according to whichvarious combinations of histone modifications can act in concert to determine the activity at a particular genomic locus (50-52). According to this, combinations of specific histone markings can act when they are present on same histone tail or on different histone tails within the same nucleosome to bring about a specific output.

The hypothesis was bolstered by different studies showing that this kind of crosstalk occurred between different modifications such as ubiquitylation of H2B and methylation of histone H3 lysine 79 ,where pre ubiquitylationwas necessary for methylation of H3K79 (53,54). Another study showed that phosphorylation of serine 10 lead to decrease in trimethylation on lysine9 of histone H3. The reason for this was that the phosphorylation of serine 10 interfered with methylation reaction itself at lysine 9 on histone H3 which was catalysed by lysine methyltransferase called SUV39H1. (50). Thus, depending upon the kind of crosstalk, different outcomes are possible. However the exact analysis of different kinds of modifications comprising Histone code remains to be fully investigated.

1.3.3 Crosstalk between DNA methylation and histone modification

Generally, it is observed that both, the modification on the DNA and histones, act together and can influence each other. Studies from McCabe et al, 2009 showed that histone modifications play an important role in targeting the Dnmts for methylation of the DNA(55). In another study, Lehertz et al 2003 showed that there exists an important link between H3K9 methylation and DNA methylation with the H3K9 methylation being an important step in recruitment of Dnmt in the pericentromeric repeat regions (56). Also it was shown that

for maintaining the DNA methylation,certain histone modification played an important role (57).

1.3.4 Histone deacetylases and their inhibitors

Acetylation of the lysine ε-amino group is brought about by lysine acetyltransferases (KATs) whereas the opposing activity is shown by histone deacetylases (HDACs). The acetylation of histones is known to lead to an open chromatin structure and to active gene transcription. On the contrary, deacetylation is usually related with gene silencing. HDACs fall into 4 different classes according to their homology to their yeast counterpart(58-60). Although the most commonly studied are from class I and II. Class III has recently gained importance with implied roles in aging .A summary of various classes and their members are summarized in Table 2.

HDAC class	Location	Members	Cofactors
Class I	Present in nucleus	HDAC1,HDAC2,HDAC3 and HDAC8	Zn++ dependent activity
Class II	Shuttle between nucleus and cytoplasm,	HDAC4,HDAC5,HDAC6, HDAC7, HDAC9 and HDAC10	Zn++ dependent activity
Class III	Nucleus	They contain seven sirtuins	NAD+ cofactor dependent activity
Class IV	Nucleus/cytoplasm	HDAC11	Zn++ dependent activity

Table2: Summary of various different classes of HDACs and the their cofactors involved in the deacetylation reaction

Most of the HDACs and KATs do not bind directly to a particular DNA motif but interact through multiprotein complexes involving corepressors and coactivators(61). HDAC1 and HDAC2 for example form homo- and heterodimers in multiprotein complexes such as those

involving Sin 3A or B or CoREST which are involved in transcriptional repression or with chromatin remodelers such as NuRD(58,59).

1.3.5 Action of Valproic acid (VPA) and sodium butyrate

To modulate the histone acetylation HDAC inhibitors (HDACi) have been identified. Most of the HDACi used today have Zn^{++} chelating groups owing to which they can fit into the active pockets of HDAC from class I, II and IV and disrupt the formation of multimeric protein complexes(62).

The HDAC inhibitors VPA and sodium butyrate are mild inhibitors of HDAC class1 enzymes (63,64). These were shown to have highest affinity for the CoREST complexes followed by NuRD and least for Sin3 mediated HDAC inhibitions(65,66).

1.4 Transgene silencing and Position effect

There are different ways a gene can be silenced. This can be attained at the level of transcription called transcriptional gene silencing (TGS) and at the level of post transcription, called posttranscriptional gene silencing (PTGS). As the thesis focuses primarily on the transcriptional gene silencing, it will be discussed in more detail. Transcriptional gene silencing is achieved by the changes in the chromatin states mediated by the alteration in the levels of DNA methylation and the dynamics of histone modification.

Most of our understanding of the epigenetic role play in differential gene expression is based on cellular genes in their natural environment. In contrast, much less is known about the crosstalk of heterologous/synthetic/transgenic regulatory elements once integrated into a specific chromosomal domain as it occurs upon genetic modification of cells. Indeed, once transgenes are integrated into the chromosomal DNA, a high clone to clone variation of transgene expression is observed. This is generally described as the chromosomal position effect(67-71). The ‘position effect’ comprises the influence of genetic elements such as enhancers which facilitate binding of transcription factors and thereby modulate the incoming promoter. In addition, it is conveyed by epigenetic modification of the incoming cassette. Indeed, these epigenetic effects have been shown to contribute to gene silencing

in vivo(72)and in vitro(2,73). This variation of expression is a concern for biotechnology whereuniform and stable transgene expression is required (74) but also when generating transgenic animals(75).

The focus of the current thesis is mainly on the epigenetic modulation affecting the transgene in a chromosomal locus. In the centre of this thesis is the Cytomegalovirus (CMV) promoter as a frequently used, highly potent promoter for driving the transgene. In addition, the thesis concerns induced expression by employing a Tetracycline inducible promoter based on the CMV promoter.

1.5 The CMV and the Tetracycline inducible promoter elements

1.5.1CMV promoter

The human immediate-early CMV (Cytomegalovirus) promoter is one of the most commonly used promoter elements for transient and stable transgene expression in plasmids as well as in viral vectors (76-78). The CMV promoter has been shown to confer ubiquitous expression and to display less tissue specificity(79).

1.5.2 Synthetic Inducible promoter (Tetracycline inducible promoter)

Based on a minimal CMV promoter that has only the TATA box for initiation of transcription, the Tetracycline induced promoter was developed. The Tet promoter is the most commonly and widely used inducible promoter system.The Tet system is composed of two elements: the Tet promoter and its transactivator proteins.The originally described tet promoter is composed of the minimal hCMV immediate early promoter (80) and a seven tet operator repeat sequences from the *E. coli* tetracycline operon (80). These so-called tetracycline responsive elements (TRE) are the binding sites for the transactivator protein (tTA or rtTA) which leads to strong transcription activation. The Tet on system is based on the ability of the rtTA transactivator protein to bind to the operator elements only in presence of the Dox (81).

Both unidirectional and bidirectional Tet promoters have been developed. The unidirectional Tet promoter has one minimal hCMV region fused to the operator regions (80,81). In addition, a bidirectional Tet promoter has been developed which had two minimal hCMV regions flanking the 7 operator repeats regions(82). The bidirectional promoter has several advantages such as: 1) it allows for co-expression of two gene products in stoichiometric amounts, 2) a reporter can be combined with non-easily assayable gene of interest and can be monitored with the help of reporter(83).

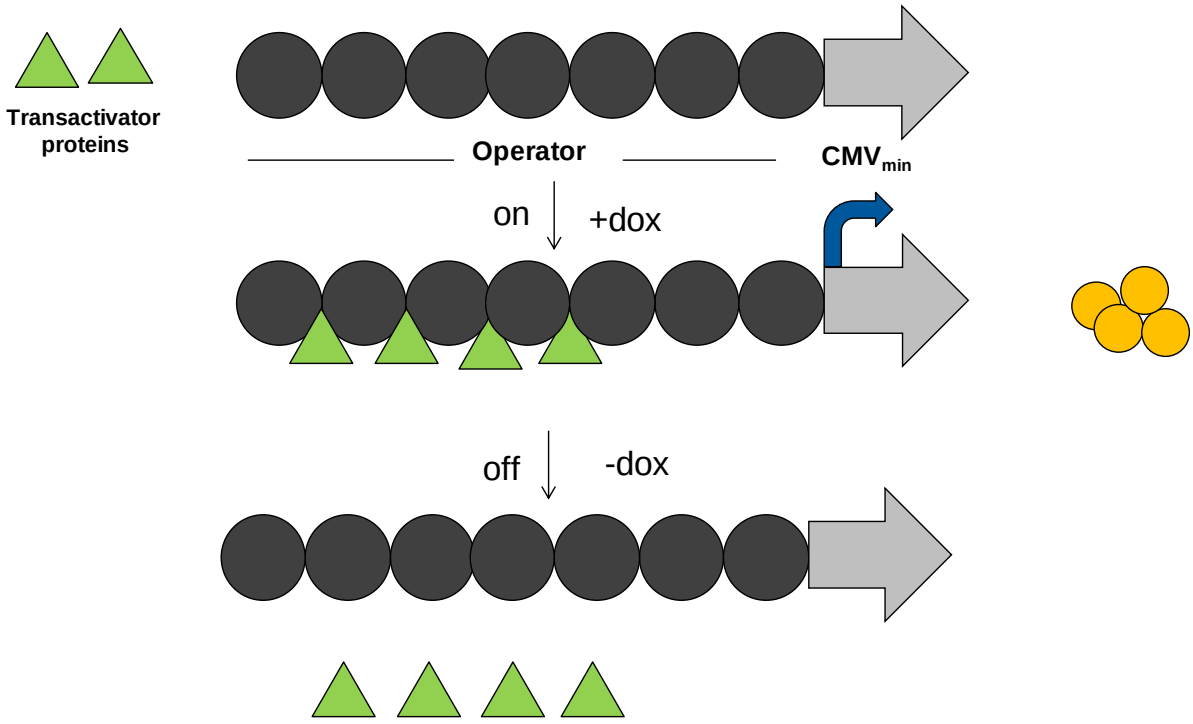


Figure 3: The Tet on system. A) The Tet-on system comprises of two components i.e thertTA transactivator and the synthetic promoter. In this system, the administration of Doxycycline renders the transactivator proteins binding. Thereby, Tet based transcription is activated. Transcription is stopped by removing the Doxycycline from the medium.

1.5.3 Utility of inducible promoter system (Tet system)

The temporal control of gene expression can be useful in cases where the experimental studies involve the expression of toxic and xenobiotic compounds or the study of genes that are embryonic lethal thus are difficult to study (84,85). The Tet system has also been used to look for the disease progression with expression of certain mutant or viral proteins (86,87), and thus it provides the researcher with a great tool to investigate questions that can be difficult to answer in normal experimental settings

1.5.4 Optimization of Tet promoter

Since its synthesis by Gossen and Bujard(80) several studies have been aimed at optimizing it further. Most of these are based on the manipulation of the transactivator proteins. These included codon optimization (88-90), introduction of nuclear localization signal (NLS)(91,92) and removal of potential splice sites(89). Another important factor was the basal expression from the Tet promoter.The basal expression of the Tet promoter might arise from various different transcription factor binding sites that are present in the Tet promoter (93). These include ISRE motifs located between operator (tetO), the GATA factors (WGATAR, whereWis A or T andR is A or G) that exactly overlie the consensussequence for the tetR (TCCCTATCAGTGATAGAGA)(93). Additionally,the murine (m) CMV element ofthe Ptethas AP2(94), cAMP-response element 1, xenobioticresponseelement 1, GCF, IFN-g, and LFA-1.In an effort to reduce the basal expression in non induced state, the consensus sequences in the TATA box and TFII binding sites were mutated and also the 5'UTR region was truncated(83).

1.6 Transgene integration sites within the chromosome

The chromosomal integration sites for transgene expression have been a topic of immense interest. The possibility of the presence of a site that supports high transgene expression "the so called genomic hotspot"and its great economic value has lead to a great expectation and extensive search for such a site. These hotspots can be used for high, stable and predictable transgene expression (95).

In the murine genome few genomic locations have been reported to support ubiquitous and stable expression of the synthetic cassette (96,97). Among these, the Rosa26 has been the most commonly used for mouse transgenesis(97)

1.6.1 A well known integration site: the Rosa26 locus

The well-known Rosa26 locus on chromosome 6 was first found in a gene trap experiment in mouse ES cells(97). It has been targeted since and has been used in a number of labs around the world to generate transgenic mice upon integration of transgenes by homologous recombination (98-100).The major advantage of the Rosa locus is its ubiquitous expression through different developmental stages.

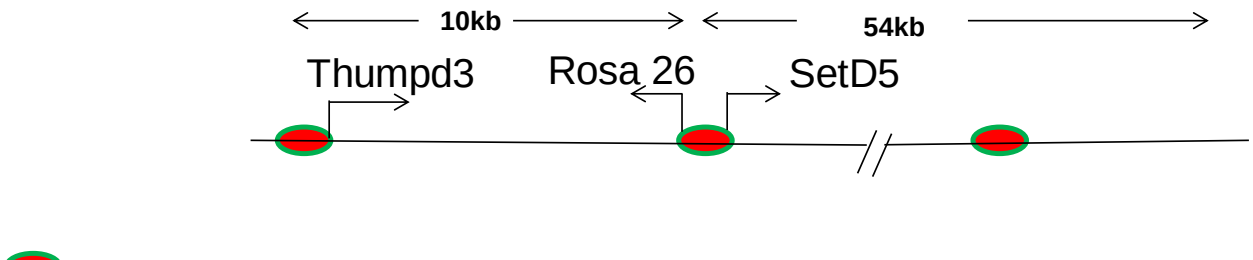


Figure 4: The Rosa26 locus (not to scale). The Rosa26 locus is present on chromosome number 6 in the mouse genome. The Rosa26 promoter is a bidirectional promoter harbouring a CpG island of roughly of about 1.5kb. Two transcripts are generated, the Rosa transcript and the Setd5 transcript. The Thumpd3 gene is present with another classical CpG island promoter about with a distance of 10kb from the Rosa promoter.

The utility of this locus over last decade has triggered the search for the corresponding Rosa26 locus in other genomes such as in rats(101), humans(102). The technological advancement has also lead to generation of targeting tools such as zinc fingers that can be used to directly target the Rosa26 locus (103,104)

1.7 Targeted integration through recombinase mediated cassette exchange (RMCE)

To deal with the heterogeneity of gene expression upon random integration of cassettes into the host's chromosomes, standard protocols employ screening procedures to identify those integration events that meet the requirements for transgene expression (e.g. stable or regulated expression). This imposes a limitation, in particular when resources for screening are restricted and/or if the monitoring of transgene expression is not straightforward. Thus, efforts have been undertaken to develop novel strategies for predictable, reliable, and, most importantly, effective manipulation of mammalian cells. The recently emerged methods are designed to target expression cassettes to preselected chromosomal loci, in particular chromosomal sites which support high level of expression (so called 'hot spots'). Following such strategies, the unpredictability associated with the random integration of expression cassettes is overcome or at least reduced since the modulation mediated by the particular chromosomal site is known. In particular, when repeatedly the same integration site is used the requirement for large scale screens for every new transgene that needs to be expressed is obviated.

Site specific recombinases were exploited for highly efficient modification of chromosomal loci in mammalian cells. The two major family members of tyrosine recombinases are Cre and Flp. Cre(causes recombination) recombinase was derived from the P1 bacteriophage. The 34bp recombination target (RT) site for Cre was named loxP (locus of crossing (x) over, P1). The other tyrosine recombinase, Flp, was identified in *Saccharomyces Cerevisiae*. Flp recombinase recognizes and binds to their sites, called the FRT sites, comprising three repeat regions flanking the unique spacer sequence(105).In a first step, a single RT or a set of specific RT(s) are integrated into the genome of a cell line (so-called tagging). This can be achieved by transfection or infection of an expression vector that carries the RT(s). Upon random integration of this vector and screening, cell clones are isolated that provide the appropriate expression pattern (e.g. high, stable or regulated transgene expression). Alternatively, the RT site(s) can be specifically integrated into defined chromosomal loci by homologous recombination. In a second step, such cell clones are then targeted with a gene

of interest by simple co-transfection of a vector harboring the RT(s) and a vector encoding the recombinase. As a result of the site specific recombination, integration of the cassette is achieved ('targeting')(106-108).

1.7.1 Recombinase mediated cassette exchange (RMCE)

In recombinase mediated cassette exchange (RMCE) targeting is a "cut and paste" mechanism that involves recombinase mediated excision of a DNA segment flanked by a set of two heterologous RTs and integration of a cassette flanked with the very same set of RTs. Thereby, an exchange of the intervening sequences is achieved. A prerequisite for RMCE is a set of heterologous RTs that cannot recombine with each other (Fig. 5)(105).

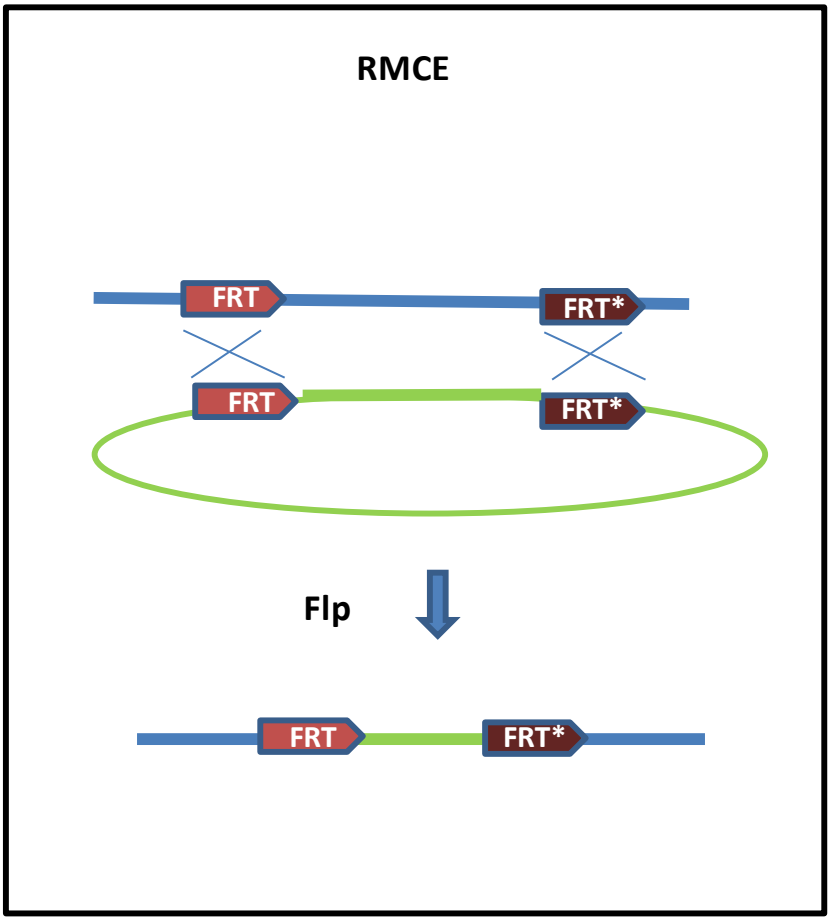


Figure 5: RMCE: Two heterospecific, non-interacting RT sites in the genome will recombine with identical RT sites delivered by an external plasmid resulting in recombinase mediated cassette exchange in presence of Flp recombinase. The reverse reaction is kinetically and thermodynamically unfavoured and virtually excluded (Figure from Kruse et al.(105))

With simple easy to use protocol, the use of recombinase based genome modification has a popular method of choice over the last years It has been used in various different fields in basic science and also biotechnology including the generation of recombinant protein expressing cell lines and also in the generation of a wide variety of transgenic mice (e.g.(99)). Basically, the concept allows to screen integration sites using reporters that facilitate easy detection (e.g. GFP) and to establish these clones as master cell lines for targeting cassettes of choice. Due to the fact that targeting is fast, efficient and precise and the genetic manipulation is defined, the expression properties of the final producer clones are predictable(105,108).

Meanwhile, there are a number of examples that exemplifies the efficient generation of cell lines using thetargeted integration of expression cassettes that express a gene or vector of interest. However, some studies also pointed out that the exploitation of the potential of defined integration sites critically depends on the careful design of the incoming targeting vectors(105).The following example might illustrate this.

A study by Gama-Norton et al. looked into the impact of the orientation of the targeting cassette in two distinct loci in 293 cells that were identified because of their capability to support high expression of retroviral vectors. They observed that upon targeting three different cassettes in sense and antisense orientation using RMCE, one of the loci could support expression of either orientation and in the other locus the high expression levels were only achieved when the targeting cassettes were integrated in the same orientation as the initial tagging vector while expression dropped up to 100 fold for the other orientation (105,109).

The RMCE based strategy can also be applied to express proteins other proteins.The work from Schucht and coworkers (110)to express G protein coupled receptors (GPCRs) from a master cell lines developed using this technology. Also, Wilke et al. used RMCE in mutant CHO Lec3.2.8.1 cells to produce glycoproteins with the well-established glycosylation pattern in a homogenous form (111).

Highlighting the major application of the RMCE technology was the study from Wiberg et al who produced human polyclonal anti-RhD antibody in CHO cells (105,112). Individual antibody expression cassettes were targeted into the same chromosomal site and these cell

clones expressing individual antibodies were mixed to generate cell pools. These pools gave rise to the production of a polyclonal antibody in a highly reproducible manner allowing a direct industrial application.Similar observations were obtained for various promoters integrated into the Rosa26 locus of mouse embryonic stem (ES) cells (105,113).

Together, these studies indicate that high expression potential of a chromosomal site can be exploited fairly nicely by means of targeted integration; however this high expression potential of chromosomal loci is also partlydependent on the cassette design. The sites favouring high, stable, and homogenous expression of transgene is highly desirable for expression of not only biotechnologically relevant proteins but also in basic research that rely on using transgenic models for investigation important bio-medically relevant questions. However, the underlying mechanism of crosstalk that occurs between an integration site and the incoming transgene has not been investigated. And although the RMCE based method allows for the repeated use of an integration site however how different transgenes are impacted epigenetically in a particular integration site remains unclear.

1.8 Aim of the study

Heterogeneity in transgene expression is frequently observed upon genetic modification of cells in basic research and biotechnology. The variable transgene expression is considered to be a result of the crosstalk of the incoming promoter cassette with cis-acting elements associated with the chromosomal site of transgene integration (position effect). Targeted integration of the transgene into the open chromatin limits this variability caused due to the chromosomal position effects thereby rendering higher predictability in transgene expression. However, this does not shield the integrated transgene from the epigenetic influences.

The objective of this study was to have a mechanistic understanding of the nature of interaction that occurs between the transgene cassettes when integrated into defined chromosomal sites. In particular, the underlying epigenetic mechanism like DNA methylation and histone modifications was to be elucidated. This concerned on the one hand the modulation of expression upon integration of CMV promoter driven transgenes into a specific site of integration in the HE293T and CHOK1 cells. Also, the epigenetic crosstalk was evaluated upon integration of synthetic Tet driven construct in the well known Rosa26 locus in transgenic mice.

2 MATERIAL AND METHODS

2.1 Materials

2.1.1 Equipment

Instruments	Manufacturer
MassARRAY Compact System	Sequenom, Hamburg, Germany
MassARRAY MATRIX Liquid Handler	Sequenom, Hamburg, Germany
MassARRAYPhusio chip module	Sequenom, Hamburg, Germany
Megafuge 3,0 R	Heraeus, Osterode, Germany
Microarray hybridisation chambers SureHyb	Agilent Technologies, Böblingen, Germany
Microarray scanner; 5 micron resolution	Agilent Technologies, Böblingen, Germany
Microarray slide holder	Agilent Technologies, Böblingen, Germany
Microscopes	Zeiss, Jena, Germany
Multifuge 3S-R	Heraeus, Osterode, Germany
MultipipettorMultipette plus	Eppendorf, Hamburg, Germany
NanoDrop	PeqLab, Erlangen, Germany
PCR-Thermocycler PTC-200	MJ-Research/Biometra, Oldendorf, Germany
PCR-ThermocyclerVeriti 384 well	Applied Biosystems, Foster City, USA
-20° C freezer	Liebherr
- 80°C freezer	Thermo Forma
4 °C refrigerator	Liebherr
BLI system	Xenogen IVIS system, Caliper
Cell Counter	Guava EasyCount, Millipore
CO2 Incubators for Cell Cultures	Labotect
Cooling Centrifuges	HettichRontana/S
	SorvallSuperspeed
	MinifugeHeraeus-Christ
	HeraeusBiofuge fresco
	Inflexible rotors
	Swing rotor
Dionized Water Supply	Millipore MilliQ
Flow Cytometer	FACS Calibur, BD Biosciences
Gel Electrophoresis Chambers	Gibco BRL Horizontal Gel Electrophoresis Apparatus
Micropipettes	Gilson
Microwave	Whirlpool
Light Microscopes	Leica Labovert FS, Nikon TMS
Epifluorescent Microscopes	AxioVision, Carl Zeiss
PCR Machine	T3 Thermocycler, Biometra
pH meter	Beckman
Photometer	Nanodrop Spectrophotometer, Peqlab

Power Supplies	Gibco BRL
	Phero-stab, Bachofer
	Biorad Power Pac
	Desaga Mains Power supply unit
Precision Weighing	Sartorius
Shaker	Heidolph
Sterile Work Benches	Steril Gard Class II Type A/B3, Baker Company Hersafe, Heraus
Table Top Centrifuges	Eppendorf
	HeraeusBiofuge
	Heraeus Christ Minifuge GL
	HettichRontana/S
Thermomixer	Eppendorf
U.V Chamber	Hanau
Vortex	Scientific Industries Vortex Genie 2
Water Bath	GFL

Table3:Summary of equipments used

2.1.2 Consumables

Material	Supplier
Bacterial Petri dishes	Nunc
Cell culture plates (96 well, 48 wells, 24 wells, 12 wells, 6 wells)	Nunc, Corning
CombiTips (0.5ml, 1.25ml, 2.0ml, 2.5ml, 5 ml)	Eppendorf
Cryogenic Vials	Corning
F/AIR Filters	Rothacher Medicals
Falcon tubes (15, 50 ml)	Greiner bio-one
PCR tubes	Biozym
Pipette tips (20µl, 200µl, 1000µl)	Star Labs
Safe Lock Tubes (1.5ml, 2.0ml)	Eppendorf
Syringe Filters (0.2µM and 0.45µM)	Sartorius
Syringes (1 ml, 5 ml, 20 ml)	Omnifix®
Tissue Culture dishes	Corning, Greiner Bio-one
MicroAmp® Optical 96-Well Reaction Plate	Life Technologies

Table 4:Summary of consumables

2.1.3 Chemicals& Kits

.

Chemicals	Supplier	Catalogue nr.
Azacytidine	Sigma Aldrich	A2385-100MG
Doxycycline hyclateBioChemica	AppliChem GmbH	A29951,0025
Decitabine	Sigma Aldrich	A3656-10MG
IsoFlo®	Allbrecht GmbH	701-005-301
10 mMdNTP Mix	Invitrogen	18427-013
2x Red PCR Master Mix	p.j.k.	302004
Sodium butyrate	Sigma Aldrich	B5887-250MG
Valproic acid	Sigma Aldrich	P4543-25G
EZ DNA methylation	Zymo research	D5002
lipofectamine 2000	Lifetechnologies	11668019
Mini prep Kit (250)	Qiagen	27106
Genomic DNA isolation	Qiagen	51306
SYBR® Green PCR Master Mix	Life Technologies	4309155
Anti H3K4 acetylation	Millipore	07-539
Anti H3K27 trimethylation	Millipore	07-449
Zero blunt cloning kit	Life Technologies	K2700-20

Table5: Summary of chemical reagents and kits used

2.2 METHODS

2.2.1 Agarose gel electrophoresis

For mixing a 1% gel, 1g of agarose was dissolved in 100ml 1 x TAE (40mM Tris/acetate pH 7.5,20mM NaOAc, 1mM EDTA) by boiling in a microwave. After relative cooling 1µl of ethidiumbromide (10mg/ml) was added and mixed while casting onto gel tray. After the gel solidified itwas transferred to an electrophoresis chamber filled with 1 x TAE. DNA samples were mixedwith 5 x loading buffer (15% Ficoll, 50mM EDTA, 1xTAE, 0.05% Bromophenol blue, 0.05%Xylenecyanole) and loaded to the gel. For the determination of the size of DNA fragments, amarker was loaded in parallel (Hyperladder I, Bioline). Electrophoresis was performed at 80-100V and 30mA. Gels were examined under UV-light (360nm).

2.2.2 Restriction analysis

For digesting DNA, restriction endonucleases were used following the reaction conditions(buffers and temperatures) as specified by the manufacturer (NEB).

2.2.3 Mammalian cell culture and mice

Cells used and culture conditions

HEK293T cells (ATCC CRL-11268) and CHOK1cells (ATCC CCL61) were cultivated at 37°C in a humidified atmosphere with 5% CO_2. The human cell line was maintained in DMEM (GIBCO, Carlsbad, CA). For the CHOK1 cell line we used DMEM medium and Ham´s-F12 medium (Invitrogen) in a ratio of 1:1. Culture media were supplemented with 10% fetal calf serum, 2 mM L-glutamine, Penicillin (10 U/ml) and streptomycin sulfate (100 µg/ml).

Targeted murine embryonic stems cells obtained from U. Hillebrand done with pEM-rTA2luc3eGFP vector (99) and cultured in the standard cell culture medium used for murine embryonic stem cells(mES cells) was composed of Knockout^{TM} DMEM optimized for ES cells (GIBCO) basal medium supplemented with 15% Heat-Inactivated FBS (SIGMA), 1x Glutamax,

1x Na Pyruvate, 1x non-essential amino acids, 1x Pen/Strep, ß-mercaptoethanol, and leukemia inhibitory factor (LIF).

Mice

Targeted murine embryonic stems cellswith pEM-rTA2luc3eGFP vector (99) were used to generate transgenic mice. This mouse line was called RosaGFPand was maintained in the breeding facilities of Helmholtz Centre for Infection Research (HZI, Braunschweig, Germany). All the experiments were carried out in accordance with the ethical board approval and guidelines. The mice were housed in isolated ventilated cages in the mouse house facility of the institute.

Vectors used in the study

Newly cloned Vectors

1) **pTAG CMVGFP(pLVsspwv)**

The lentiviral self-inactivating tagging vector (pTAG CMV GFP) was derived from pJSARGFP(114). The lentiviral SIN backbone of pJSARGFP (ClaI-NheI fragment) comprising the LTRs, REV responsive element and woodchuck hepatitis regulatory element (WPRE) was ligated to a reporter cassette comprising the CMV promoter driving eGFP reporter gene. This cassette was further flanked by a set of heterospecific, non-interacting Flp recombination-target sites FRT-WT and FRT-F5.

2) **pEM-OCGrtTA2luc3eGFP(CpG free BiTet with dual marker-Luciferase and eGFP)**

The CpG free bidirectional Tet promoter was synthesized from GeneArt with Xho1 and ECOR1 restriction sites which were used to clone this CpG free promoter fragment into the pEM-rtTA2luc3eGFP vector to obtain a CpG free bitet promoter targeting vector.

Summary of all the vectors along with their internal database reference bacterial number(B.Nr) and DNA number(DNA.Nr) is summarized in the table 3 below.

S.NO	Name	B.Nr	Source or references
1	pEM-rtTA2luc3eGFP	B6272	(97).
2	pCMVRTA2HYG	B6321	This study
3	pTAR CMV-RFP	B6844	(115).
4	PTAR SV40-RFP	B6589	(115).
5	pTAG CMV GFP(pZLE-1))	B7023	This study
6	pEM-OCGrtTA2luc3eGFP	B7713	This study

Table 6: Summary of all the vectors used in the present study.

2.2.4 Lentiviral gene transfer

HEK293T cells were used for lentivirus production as specified in Norton et al (109). Briefly, HEK293T cells were transfected using the calcium phosphate method with four different plasmid constructs: the tagging vector (pTAG CMV GFP), envelope-encoding plasmid (pLP-VSVG), gag/pol helper plasmid (pLP1) and REV expressing plasmid (pLP2). After 12 hours the medium was changed.48 hours after transfection the supernatant containing the lentiviral particles was harvested and filtered through a 45 µm filter. The virus supernatant was titered by infecting 293T cells with serial dilutions and subsequent flow cytometric evaluation. For generation of single copy tagged clones, 1 x 10^5 HEK293T or CHOK1cells were seeded and infected with 1000 viruses (MOI 0.01) in the presence of 8 µg/ml protamine sulphate (Sigma Aldrich). The day after infection, medium was exchanged.

2.2.5 Recombinase mediated cassette exchange (RMCE)

Targeting of the tagged clones was done by Flp mediated recombinase cassette exchange according to previously published protocols (116). Briefly, the FLPe expression vector pFlpe(116) was cotransfected with the targeting vectors pTAR CMV RFP and pTAR SV40 RFP,

respectively, using lipofectamine 2000 (Life technologies) in a molar ratio of 3:1 according to manufacturer's protocol. Media was exchanged after 5 hours. 10 days after transfection, the cells were sorted for lack of GFP expression resulting from successful RMCE. Targeted vector integration was confirmed by PCR.

For mES cell targeting, 1 x 10^4were seeded in the 6 well plate without feeders and the FLPe expression vector pFlpe(116)werecotransfected with the targeting vectorspEM-rTA2luc3eGFP(99) and CpG free pEM-oCGrTA2luc3eGFP, respectively, using lipofectamine 2000 (Life technologies) in a molar ratio of 3:1 according to manufacturer's protocol(99). Briefly, media was exchanged after 5 hours. After 48 h post transfection, the transfected ES cells were transferred to feeder coated 10 cm cell-culture dishes and selection pressure with G418 at a concentration of 0.4 mg/ml was applied. As a negative control, untransfected mES cells and mES cells transfected with only the Flp recombinase expression plasmid was always included. Selection was usually carried out for 8–10 days during which it was ascertained that all the cells in the negative control were killed. Putative RMCE targeted G418 resistant subclones obtained were then picked and analyzed for GFP expression

2.2.6 Chemical treatments

For the chemical treatments 1 x 10^5 cells were seeded in a 24-well plate format. HDAC inhibitors sodium butyrate (Sigma Aldrich) and Valproic acid (Sigma Aldrich) were dissolved in PBS and added at a final molecular concentration of 1µM and 300mM, respectively.

For treatment of mES cells,1 x 10^5 cells were seeded in a 12-well plate format. Dnmt inhibitors azacytidine (Sigma Aldrich) and Decitabine (Sigma Aldrich) were dissolved in PBS and added at a final molecular concentration of 1µMand 0.5µM, respectively.

Treated and untreated control cells were cultivated for 72 hours and then harvested for further analysis

2.2.7 Flow cytometry

Flow cytometry was used for the analysis of transgene expression with BD FACScalibur. Sorting of the cells was done with FACSaria, FACScanto, using appropriate laser and filter settings for GFP and RFP, respectively. A gating strategy was used to eliminate doublets and dead cells or debris. Results were quantified with the FlowJo 7.6 software.

2.2.8 Chromatin Immunoprecipitation (ChIP)

Chromatin immunoprecipitation was done using the ChIP-IT High sensitivity Kit (Active Motif). The protocol was followed according to the manufacturer's instructions. Briefly, about 1,5 x 10^7 cells were fixed and their chromatin was sheared upon sonication. Sonicated samples were incubated at 4°C overnight with specific antibodies against H3K27me3 (Millipore, cat.no-07-449) and H3K4ac (Millipore, cat.no-07-53907-53907-539D07-539DEdddd07-539) (,6µg/reaction). Immunoprecipitation was done using Protein G agarose beads. Finally the DNA was purified after the reversal of crosslinks and subjected to quantitative PCR.

For quantitative PCR the following primers were used for the CMV promoter region (forward primer 5'-AAGTACGCCCCCTATTGACG-3' and reverse primer 5'-AAACCGCTATCCACGCCCAT-3'. PCR conditions were as follows: 10µl SYBR green RT-PCR mix (Qiagen), 1µl (10mM) forward primer, 1µl (10mM) reverse primer, 8µl immunoprecipitated DNA. Real-time PCR was performed on a LightCycler 480 apparatus (Roche). The reagents, primers and samples were added in a 96-well-plate (Roche). All assays were performed in triplicates. The reaction was performed according to the following conditions: 1. Pre incubation: 95°C 15min. 2. Amplification: a. Denaturation 95°C for 15 sec. b. Annealing 58°C 20 sec. c. Elongation 72°C 30 sec (45 cycles). 3. Melting curve a.95°C 5 sec.b.70°C 1min c.95°C continuously.4. Cooling 40°C 30 sec.

Percentage input was used to quantify the enrichment.With this method, signals obtained from the ChIP are divided by signals obtained from an input sample prepared separately according to the manufacturer protocol. One percent of the input was taken(1:100 diluted) and calculated accordingly with final adjustment of 6.64 being subtracted from the 1% input

Ct value (if the starting input fraction is 1%, then a dilution factor (DF) of 100 or 6.644 cycles (i.e., log2 of 100) is subtracted from the Ct value of diluted input).

2.2.9 Bisulfite Sequencing

A) Determination of methylated cytosines using Epitypher(adapted from the lab of Prof. Dr. M Rehli)

A common method for analyzing cytosine methylation is bisulfite conversion of DNA followed by sequencing. Cytosine-derivates undergo reversible reactions with bisulfite yielding a 5,6-Dihydro-6-sulfonate, which deaminates spontaneously. After that the sulfate is eliminated under alkaline conditions, leaving uracil. 5'-methyl cytosine is not affected by this reaction and so unmethylatedcytosine appears as a Uracil in the sequencing reaction whereas 5'-methyl cytosine remains cytosine.

In the following procedure this methylation specific difference is used for generating methylation depending mass differences that areanalysed by mass spectrometry. All reagents in this section were obtained from the EZ DNA methylation kit (Zymo). 1 µg of genomic DNA was brought to a volume of 45 µl and was diluted with 5 µl M-Dilution Buffer, mixed and incubated at 37°C for 15 minutes. After incubation 100 µl of CT Conversion Reagent was added, lightly vortexed and incubated in the lightcycler with the following protocol:

Step 1:	95°C	30 sec
Step 2:	50°C	15 min
Step 3:	Repeat steps 1-2 for 20 cycles	
Step 4:	4°C	hold

Afterwards the samples were incubated on ice for 10 minutes, 400 µl of M-Binding Buffer was added and the sample was loaded on a Zymo-Spin I Column placed in a 2 ml collection tube. DNA was bound by centrifuging at full speed for 15-30 seconds, washed with 200 µl M-Wash Buffer, centrifuged again for 15-30 seconds and then treated with 200 µl M-Desulfonation Buffer for 15 minutes at room temperature. After incubation the column was

centrifuged for 15-30 seconds, washed twice with 200 µl M-Wash Buffer centrifuging 30 seconds and 1 minute respectively at full speed to remove wash buffer residues. To elute the DNA 100 µl water was added directly to the centre of the column and centrifuged 30 seconds at 3000 rpm. The procedure yields 100 µl of bisulfite converted DNA with a concentration of 7-8 ng/µl.

PCR-amplification

Polymerase Chain Reaction (PCR) allows the specific amplification of DNA segments (see 4.2.1). The PCR-reactions were prepared in 384 well plates (ABgene) with the following reagents according to the manufacturer

Component	*Volume for single reaction*	*Final concentration*
ddH2O	1.42 µl	N/A
10x HotStarBuffer	0.5 µl	1x
dNTP mix 25 mM each	0.04 µl	200 µM
5 U/µl Hot Star Taq	0.04 ml	0.2 U
DNA Template	1 µl	5-10 ng

To each reaction 2 µl primer mix was added, giving a final reaction volume of 5 µl, with the concentration of 500 pM of the forward and reverse primer. Then the plate was sealed with AB-0558 spun down, centrifuged and incubated in a Veriti 384 well thermal cycler (Applied Biosystems) with the following programme:

Step	*Temperature*	*Time*	*Cycle*
Initial denaturation	94°C	4 min	1
Denaturation	94°C	20 sec	45
Annealing	59°C	30 sec	

Elongation	72°C	1 min	
Final elongation	72°C	3 min	1
Cooling	4°C	hold	1

Shrimp Alkaline Phosphatase (SAP) Treatment

Unincorporated nucleotides can disturb downstream applications and are therefore enzymatically inactivated. Under alkaline conditions SAP removes phosphate groups from several substrates including deoxynucleotide triphosphates, rendering it unavailable for further polymerase catalyzed reactions. The SAP solution was prepared as follows:

Component	*Volume for single reaction*
RNAse free water	1,7 µl
SAP	0.3 µl

2 µl of the SAP solution was added to each PCR-reaction with the 96 channel pippeting robot MassARRAY Liquid Handler and FusioTM Chip Module (Matrix). The plate was sealed with AB-0558, centrifuged and incubated as follows on a Veriti 384 well thermal cycler (Applied Biosystems):

Step 1:	37°C	20 min
Step 2:	85°C	5 min
Step 3:	4°C	hold

In vitro transcription and RNaseA treatment

The PCR reaction is transcribed into RNA *in vitro* with the T7 RNA polymerase, which is guided to the amplified PCR-products by the introduced T7 promoter tag in the reverse primer. The transcribed RNA is in the same reaction enzymatically cleaved by RNaseA, cleaving specifically after cytosine and thymine. T-specific cleavage is achieved by using modified cytosine triphosphate nucleotides which protect from RNaseA digestion when

incorporated in an RNA polymer. The RNase and T-cleavage mix was prepared according to the manufacturer's instruction:

Component	*Volume for single reaction*
RNase free water	3.21 µl
5x T7 Polymerase buffer	0.89 µl
Cleavage Mix (**T mix**)	0.22 µl
DTT (100 mM)	0.22 µl
T7 R&DNA Polymerase (50 U/µl)	0.4 µl
RNase A	0.06 µl
Total volume	5 µl

60 5 µl of the mix and 2µl of the SAP treated PCR reaction were transferred into a new 386-well plate with the 96 channel pippeting robot MassARRAY Liquid Handler and FusioTM Chip Module (Matrix), sealed with AB-0558, centrifuged and incubated on a Veriti 384 well thermocycler C (Applied Biosystems) for three hours at 37°C.

Desalting of cleavage reaction: resin treatment

Because salt ions are co-vaporised when acquired during MALDI-TOF analysis they are therefore visible in the mass-spectra. This would irritate the analysis of the mass-spectra and therefore the reactions need to be desalted. For desalting of the transcription/cleavage mix 20 µl water was added to each reaction with the MassARRAY Liquid Handler (Matrix) followed by the addition of 6 mg CLEAN resin per reaction. This mix was rotated for slowly for 10 minutes and spun down to collect the resin at the bottom of the wells.

Transfer on SpectroCHIP and acquisition

The SpectroCHIP holds the matrix on which the sample probes are spotted and consists of a crystallized acidic compound. When the analyte is spotted on the matrix its solvent dissolves the matrix, and when the solvent evaporates the matrix recrystallizes with analyte-molecules spread enclosed in the crystals. The DNA samples are transferred on a

SpectroCHIP with the Phusio Chip Module and analysed with the MassARRAY Compact System MALDI-TOF MS (all Sequenom). The co-crystallized analyte is acquired with a laser while the matrix is predominantly ionized, protecting the DNA from the disruptive laser beam. Eventually the charge is transferred to the sample and charged ions are created which are accelerated in a vacuum towards a detector that measures the particle's time of flight.

Data processing

Acquired data was processed with the EpiTyper Analyzer software (version 1.0.5, Sequenom). The MS is calibrated with a four point calibrant (Sequenom) containing 1479, 3004, 5044.4 and 8486.6 kDa particles. Relative to this calibration the accelerated analytes generate signal intensity (y-axis) versus mass (kDa, x-axis) plots. With the sequence of every amplicon known, the software can virtually process the sequence and predict the fragments from the in vitro transcription/RNase digestion and relocate CpG units. If fragments contain a single CpG this is called a CpG-site. More CpG-sites within one fragment are summarized to a CpG-unit get a sum methylation value since the software averages the methylation of the individual CpG-sites. If expected and incoming information match, the signal intensities of the methylated and unmethylated DNA templates are compared and quantified. A normal calibrated system is able to measure fragments between a range of 1500 and 7000 Dalton. Fragments outside of this range and fragments whose mass peaks are overlapping with multiple other fragments cannot be analysed

B) DNA methylation analysis by cloning and sequencing

All reagents were obtained from the EZ DNA methylation kit (Zymo). The protocol was followed according to the manufacturer's instructions. Bisulfite converted DNA was amplified with bisulfite primers specific for CMV promoter (forward primer 5'-GTATATGATTTTATGGGATTTTTTTATTTG-3' and reverse primer 5'-ATTCACTAAACCAACTCTACTTATATAAAC-3'). The reaction was performed on PCR machine (Biometra) according to the following conditions: 1. Pre incubation: 95°C 15min. 2.

Amplification: a. Denaturation 95°C for 30 sec. b. Annealing 55°C 60 sec. c. Elongation 72°C 60 sec (45 cycles). 3. Final elongation 72°C 7min. 4. Final hold 4°C. Amplified PCR products were integrated into the PCR blunt cloning vector (Invitrogen) using the protocol according to the manufacturer's instructions. Upon transformation, single independent clones were picked and expanded. Plasmid DNA was isolated, purified and sequenced (HZI sequencing facility). Sequences were then analysed for the presence of C to T conversions indicating unmethylatedCpGs.

2.2.10 Bioluminescence imaging with Xenogen IVIS 200 (for luciferase expression in mice)

To monitor the luciferase/bioluminescence activity of injected cells in a particular animal repeatedly without the need for sacrificing the animal the *in vivo* imaging technology (Xenogen/Caliper). Initially, a grey-scale image of the sample in the light tight chamber was automatically taken. Then, photons were collected by a sensitive CCD camera and the signals were overlaid to the grey-scale image. Parameters such as the aperture of the lenses, exposure time and binning are adjusted so as to regulate the sensitivity and resolution of the image. The field of view was adjusted to obtain images of up to 5 mice in one round. Analyses of images were performed with the Living image 2.60.1 (Igor Pro 4.09A) computer programme. Animals were first anaesthetized in the induction chamber by 2-2.5% isoflurane (Albrecht). Mice were then injected intra-peritoneal with 100µl of luciferin (30mg/ml in PBS, Synchem OHG) and placed on the heated (37°) platform in the acquisition chamber. Anaesthesia was maintained by constant administration of isoflurane via nose cones while images were taken. All the images acquired were analysed using the Living Image 2.60.1 (Igor Pro 4.09A) software.

2.2.11Luciferase expression in vitro culture

1 million cells were washed with PBS and were suspended in 500µl 1X RLB buffer (Promomega) and were frozen in -20 C for 20 minutes. These cells were then thawed and 20µl of the supernatant was mixed with 100ul of the Luciferase Assay Reagent into luminometer tubes, luminance was then quantified. Final quantification involved the basal

correction where the expression from the cell line without the luciferase was subtracted from test samples. Further, the expression was related to the total amount of protein quantified using Bradford assay.

2.2.12 Computerprograms and software analysis

This thesis was written using the Microsoft Office 2010 package specifically Microsoft Word, PowerPoint, Picture Manager and Excel. Flow cytometry results were analysed with either the CellQuest Pro program or FlowJo version 10. RT-PCR reactions were analysed either by Microsoft Excel or with the LightCycler 480 SW1.5 Software and Microsoft Excel. Graphs were generated either using Microsoft Excel or GraphPad Prism software. In vivo bioluminescence data was analyzed with the Living Image 2.60.1 software from Xenogen. Microscopic images were analysed with ImageJ software.

2.2.13 Histological sections

Mice were sacrificed and tissues from different organs were fixed for 48 hours in Formalin and then dehydrated in 70% ethanol before processing for GFP staining by the histopathology facility of the HZI. The analysis of GFP and micrograph preparation was performed by Dr. Marina Pils and Dr. KatrinSchlarmann of the Histopathology Facility of the HZI.

2.2.14 Statistical evaluation

Statistical analyses were performed using GraphPad Prism 5.0 software (GraphPad, SanDiego,CA). Data between two experimental groups were analyzed using Mann-Whitney test. The differences were considered as statistically significant if $p<0.05$(*), $p<0.01$(**), $p<0.001$(***).

3 RESULTS

The thesis comprises two different approaches to elucidate the crosstalk of transgene cassettes with the chromosomal environment of the integration site. In the first part of the Result section (3.1), a conventional screening approach relying on random transgene integration is used to identify clones with high transgene expression levels.

In the second part of the Result (3.2), the focus will shift to the specific crosstalk observed upon targeted integration of transgene cassettes into a previously identified chromosomal site.

3.1. Crosstalk of transgene expression cassettes in randomly obtained sites in HEK293T cells and CHOK1 cells

3.1.1 Tagging of genomic loci with FRT sites using lentiviral transduction

In order to identify chromosomal sites that support transgene expression, a screening cassette with GFP reporter gene driven by CMV promoter was randomly integrated into the genome of CHOK1 cells and HEK293T cells. The overall outline of the strategy is presented in Fig 6. As a reliable method for achieving single copy integrations lentiviral transduction was used. A self-inactivating (SIN) lentiviral vector with a deletion of the viral promoter in the 3'LTR was employed to avoid interference of the viral regulatory elements with the CMV promoter upon infection (Figure 6). In addition, the reporter cassette was flanked with a set of heterologous FRT sites (FRT-WT and FRT-F5) which provides the option for subsequent exchange of cassettes by Flp recombinase mediated cassette exchange (108,117). To ensure a single copy integration of the screening cassette, infection was performed at a multiplicity of infection (MOI) of 0.01. Thereby, statistically, 99% of infected cells carry a single copy integration of the expression cassette. 10 days after lentiviral infection cells showing high GFP expression (>10^3 arbitrary units, a.u.) were isolated by single cell sorting (Figure 6). Individual cell clones were established in 96 well plates and expanded for further analysis.

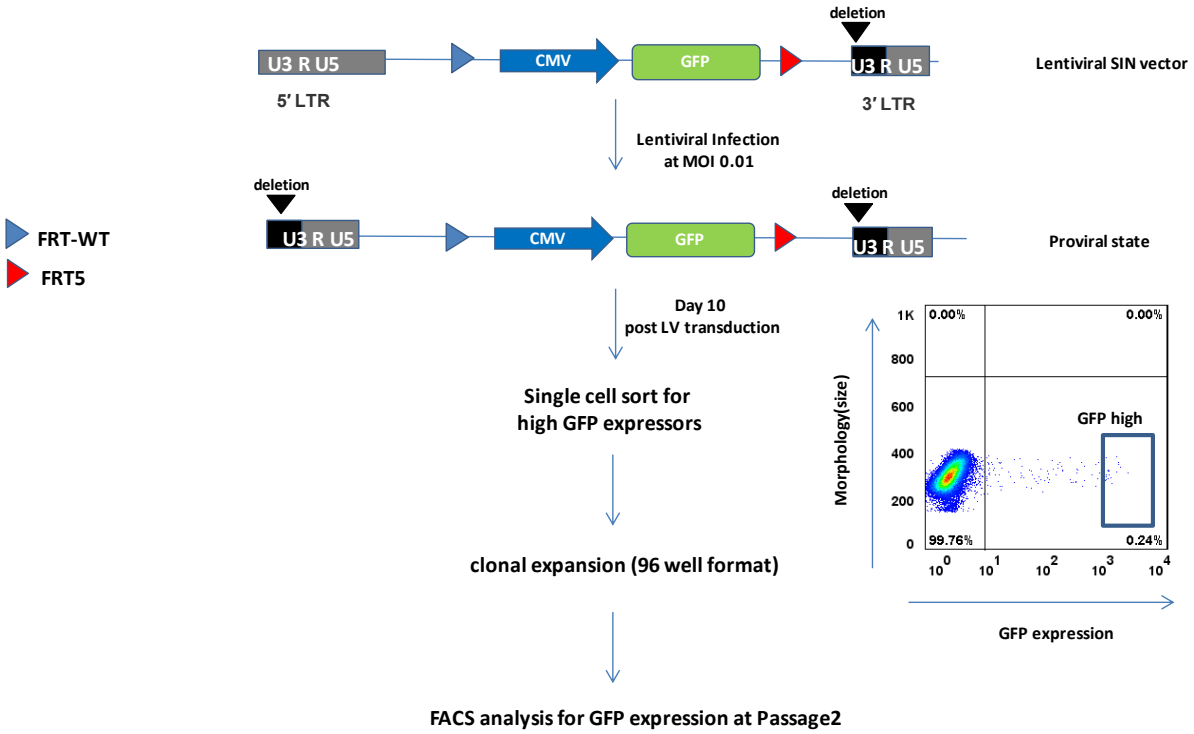


Figure 6: Strategy for generation of single copy transgene expressing clones. A lentiviral SIN vector was used to transfer a CMV driven GFP expression cassette flanked by non-interacting FRT-WT and FRT-F5 sites. The lentiviral vector and the proviral state after lentiviral infection and subsequent integration into the chromosome are depicted. The figure depicts a representative flow cytometry plot obtained of HEK293T cells at day 10 after infection at a multiplicity of infection (MOI) of 0.01. High GFP expressing cells were sorted in a single cell/well in a 96 well plates and expanded to derive clonal populations that were analysed further for GFP expression at passage 2.

3.1.2 Expression analysis of CMV-GFP in HEK293T and CHOK1 clones

The cell clones from both HEK293T and CHOK1 were established and then individual clones from the two cell lines were randomly picked and expanded. In total 55 clones from HEK293T and 60 clones from CHOK1 were analysed for GFP expression using flow cytometry. Interestingly, the analysis of the clonal population from both the cell lines showed varying GFP expression levels (Fig. 7 shows representative clones for HEK293T cells; See also Fig. XX for an overview of all the clones analyzed).

Although these clones were sorted for high GFP expression (>10^3a.u), most of the clones showed an overall decrease in the expression levels in both the cell lines. The clonal populations varied significantly in spite of arising from cells that were sorted for similar expression (>10^3a.u).

Another interesting observation was the large variance in the clonal population of both the cell lines. Most of the clones had cell fractions that showed significant reduction in expression and also loss of expression. Clones like 02, 06, 31, 42, 44 and 54 in Fig 7 showed this large variance with distinct cell populations in both expressing (>10^1 a.u) and non-expressing (<10^1a.u). The clones that showed nearly 10% cell populations losing the expression below (10^3a.u) were characterized as heterogeneous while the clones that had major percentage (>90%) of cells still expressing in excess of 10^3a.u were classified as homogeneous.

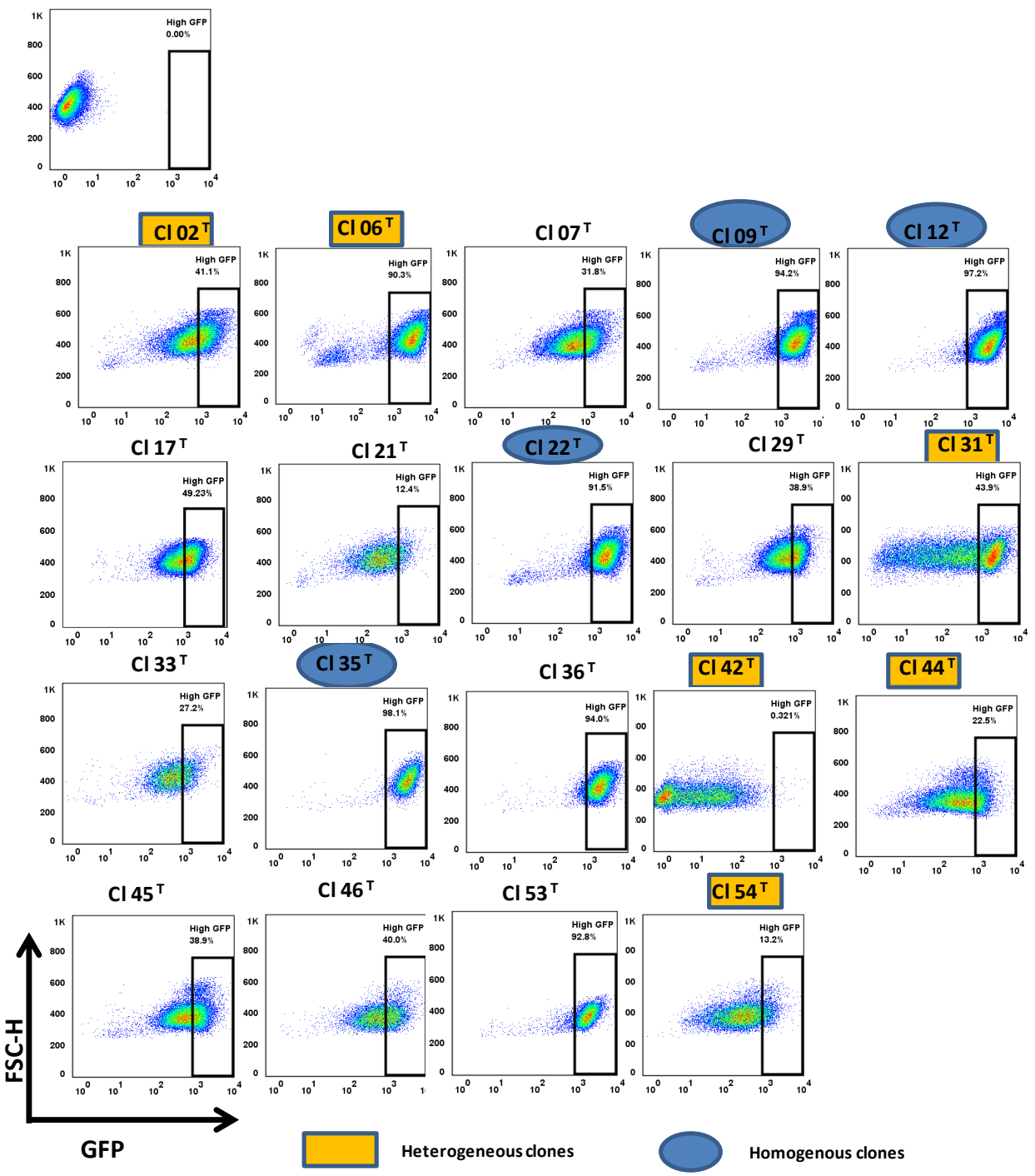


Figure 7: Expression profiles of HEK293T cell clones 2 passages after initial sortingaccording to Fig 6. FACS plots show the GFP expression profiles of indicated HEK293T clones.Clones from HEK293T with homogenous expression (>90% expressing in excess of 10^3a.u) are highlighted with blue circle and heterogeneous marked by yellow rectangle(clones where nearly 10% of cell fraction shifted below 10^3a.u).

3.1.3 Intraclonal variation in HEK293T and CHOK1 cell clones

The 55 clones from HEK293T and 60 clones from CHOK1 were categorized with respect to the mean fluorescence intensity (MFI) of the overall population as well as the percentage of cells providing high (>10^3a.u.), low (10^3 -10^1 a.u.) and no GFP expression (<10^1 a.u).

The results for HEK293T cells and CHOK1 cells are summarized in Figure 8 and 9, respectively. Although the initial sorting was restricted to cells that show >10^3a.u., in the established clones only about 10% (6/55) of the HEK293T clones showed homogenous and high GFP expression with >90% of high GFP expressing cells with an MFI > 2000 a.u. (marked by * in Figure 8).

In contrast, none of the CHOK1 clones provided homogenous and high GFP expression. This suggests cell type specific differences in the inactivation of transgene expression. Most of the CHOK1 clones (88%) and also 15% of the HEK293T cell clones (8/55) showed a significant drop of expression below an MFI of 1000 a.u. (Figure 9). In general, the overall expression level was higher in HEK293T clones in comparison to clones derived from CHOK1 cells. Three CHOK1 clones (12, 53 and 60) showed no GFP expression (Figure 9).

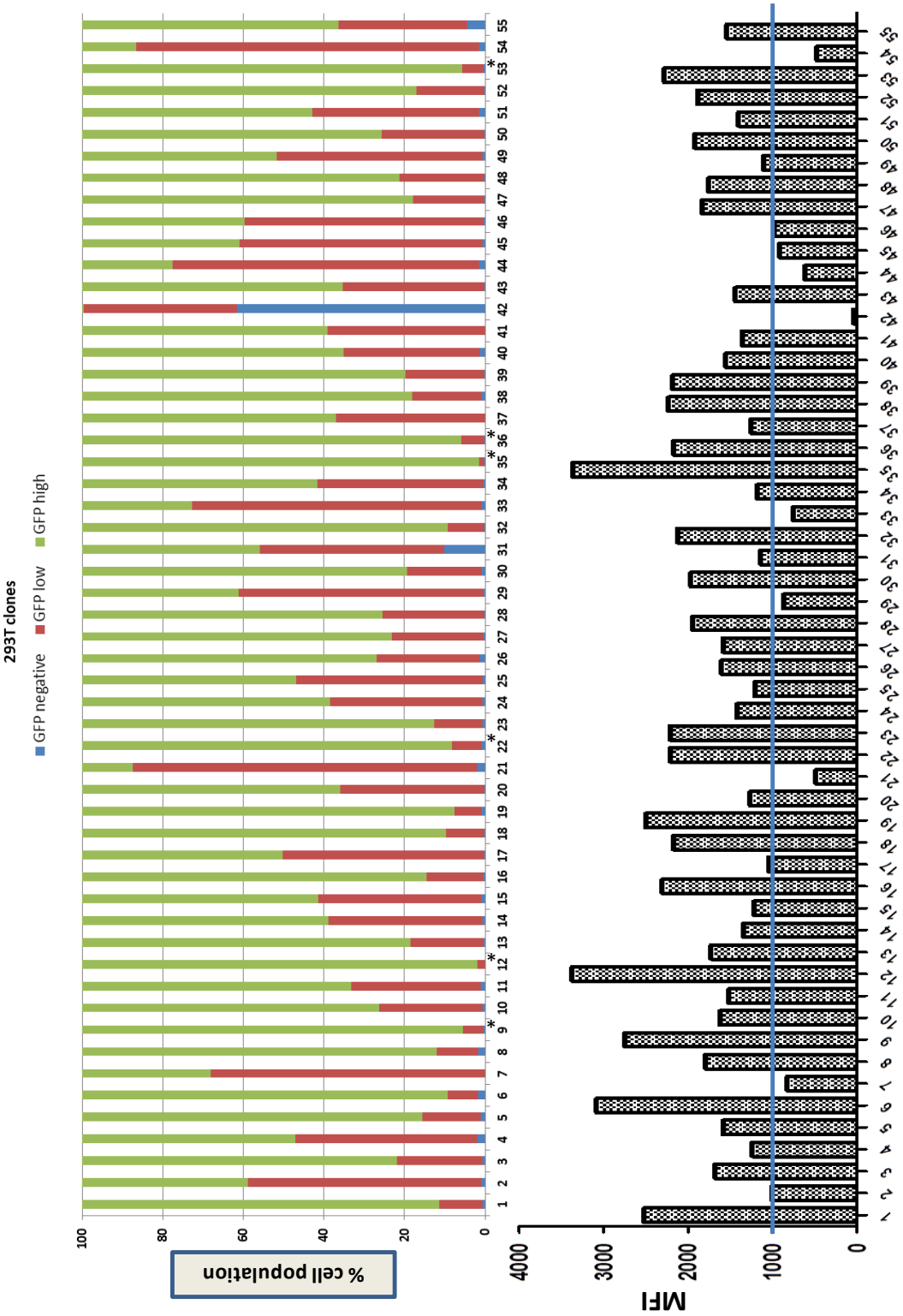


Figure 8: Summary of the GFP analysis (% expressing cell populations and mean fluorescent intensity (MFI)) in 55 randomly selected HEK293T clones. Three colours in a bar represent three populations within a clone classified according to transgene expression levels (blue <10^1, red-10^1-10^3, green >10^3). The blue line represents the MFI of 1000 a.u.**.**The HEK293T clones that showed homogenous and high GFP expression with >90% of high GFP expressing cells with an MFI > 2000a.u. (marked by *).

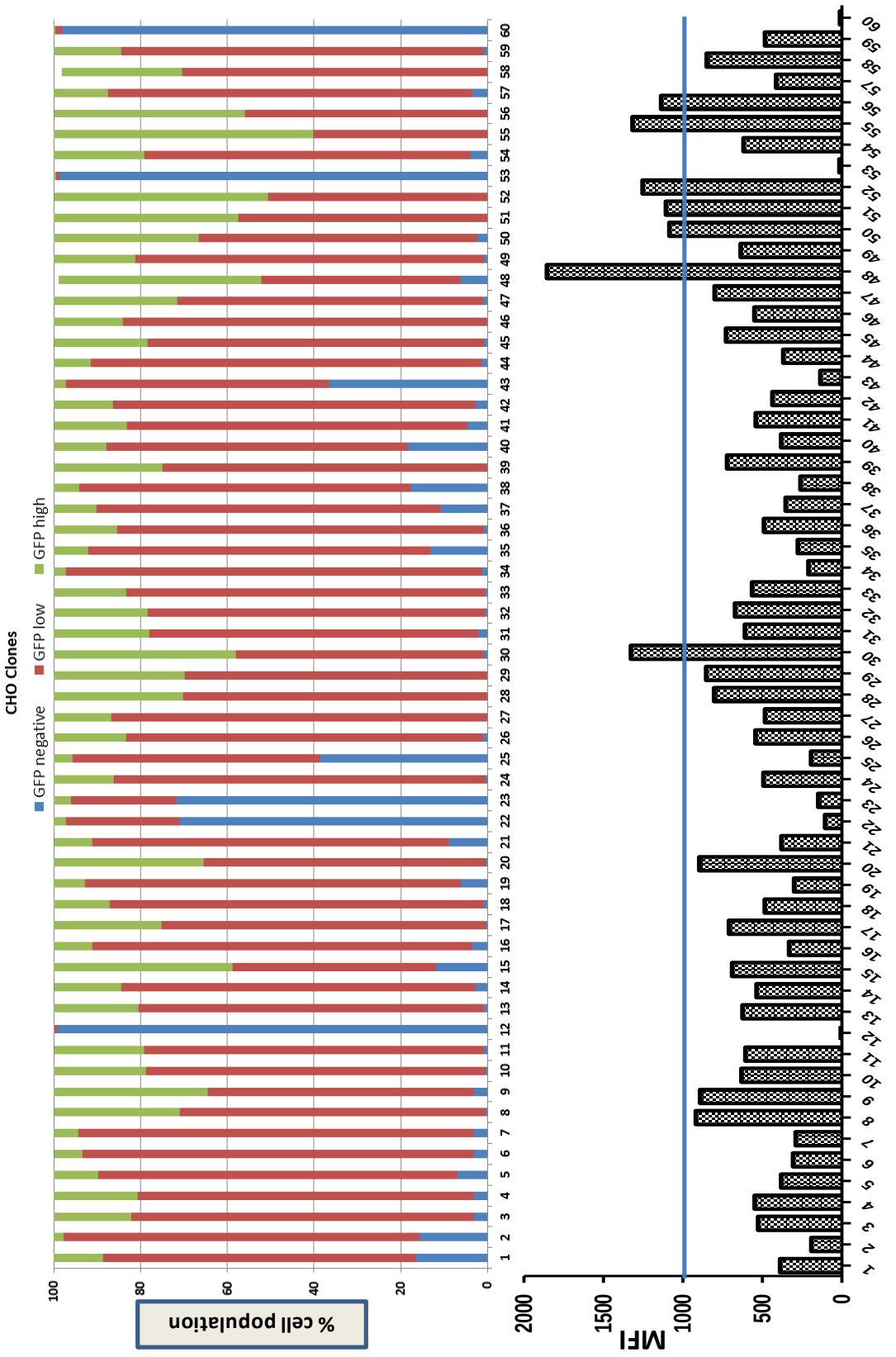


Figure 9: Summary of the GFP analysis (% expressing cell populations and mean fluorescent intensity (MFI)) in 60 randomly selected CHO clones. Three colours in a bar represent three populations within a clone classified according to transgene expression levels (blue $<10^1$, red 10^1-10^3, green $>10^3$). The blue line represents the MFI of 1000 a.u.

The heterogeneous expression of GFP suggested that transgene expression changed within the first days of cultivation/passaging. To analyze the underlying mechanism five cell clones 12^T, 17^T, 31^T, 42^T and 54^T from HEK293T cells with different patterns of heterogeneity were selected for in depth elucidation of the underlying mechanism (see Figure 7 for the specific expression pattern of these clones). While clone 12^T and 17^T showed a distinct GFP expression profile, the other three clones were characterized by highly heterogeneous expression profiles with high intraclonal variations (Figure 7). In all of the clones a small but distinct population of non-expressing cells was detected.

The GFP expressing and GFP non-expressing cells of these five clonal populations were sorted at passage 3. After sorting, the GFP expressing and non-expressing populations, called PS and NS respectively, were expanded. To investigate if this loss of GFP expression was due to loss of the transgene cassette the chromosomal DNA was isolated from the NS cell fraction of HEK293T cell clones and also from the completely silenced CHO clones (#12, 53, 60 see Figure 9). A PCR was performed using primers (P1 and P2) binding to CMV promoter region of the transgene. Clones with transgene integration were expected to give rise to a 283 bp fragment (Figure 10).

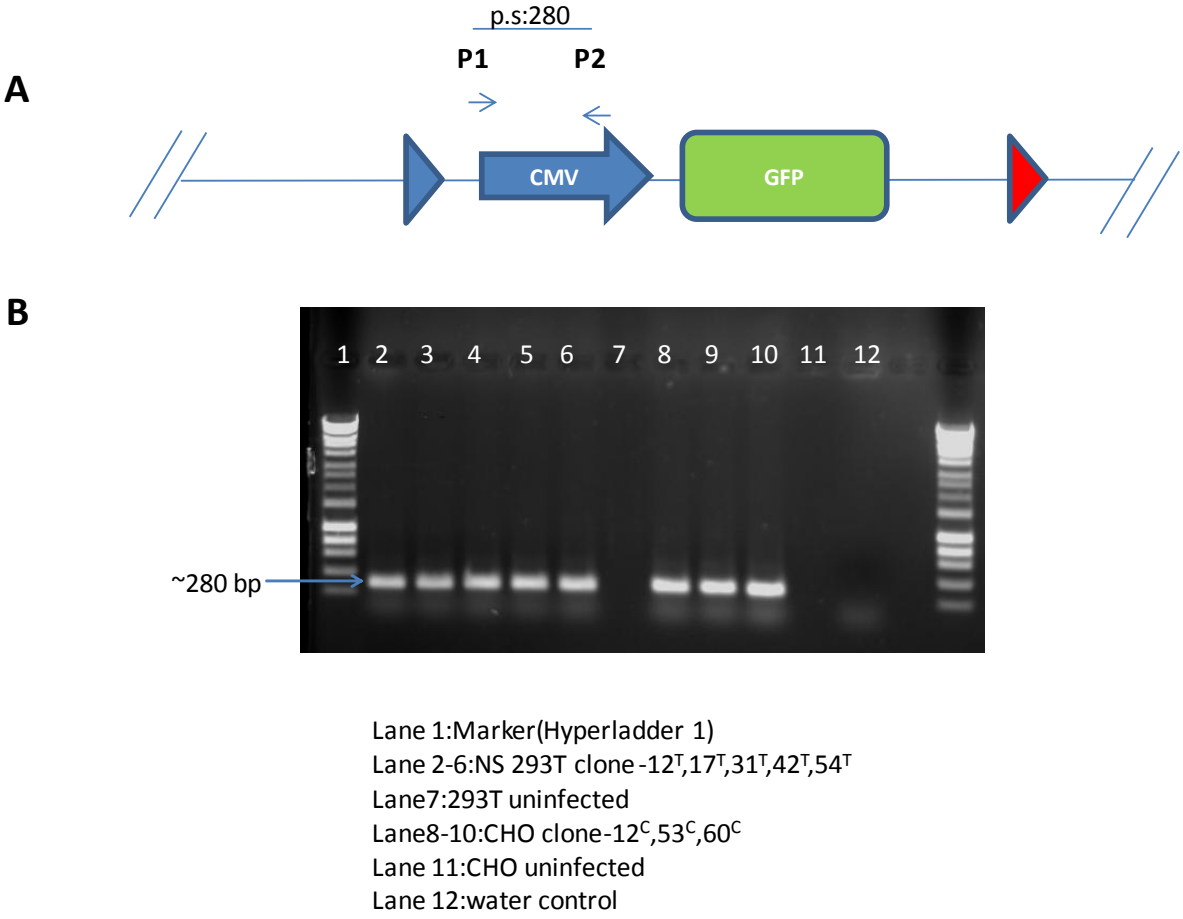


Figure10: Normal PCR for confirming the integration of transgene in the negative sorted (NS) HEK293T clonal fractions with no GFP expression and three uniformly silenced clones in CHO cells. Primer pairs P1 and P2 were used that amplified a region of 280bp in the CMV promoter region of the transgene as indicated.

Although heterogeneous, but none of the clones completely lost expression in HEK293T cell clones. This was in contrast to CHOK1 where 3 clonal populations (12^C, 53^C, 60^C see Figure9) showed complete silencing (>99% cell not expressing GFP). Also, the fractions of cells not expressing (indicated by blue region in the bar graphs of Figure8 and 9) or low expressing (indicated by red region in the bar graphs of Figure8 and 9) were more pronounced in CHOK1 derived clones. Henceforth, the statistical evaluation indicates a pronounced instability of transgene expression in CHOK1 cells.

3.1.4 Heterogeneity in expression is not correlated to CpG methylation in the CMV promoter

The CMV promoter – although providing high levels of expression in many cell types – has been shown to be sensitive to DNA methylation both in vivo and in vitro (118-120). The degree of methylation was found to correlate with the expression status of the cells (121). Thus, it was hypothesized that the heterogeneity in transgene expression might be modulated by epigenetic modification of the CMV promoter. The DNA methylation status of the promoter in the sorted NS and PS populations was analysed (at passage 3). In particular, the focus of the analysis was on a 283 bp fragment of the CMV promoter encompassing the TATA box and essential transcription factor binding sites as well as 14 CpG sites which could be potentially methylated (see Figure 11A).

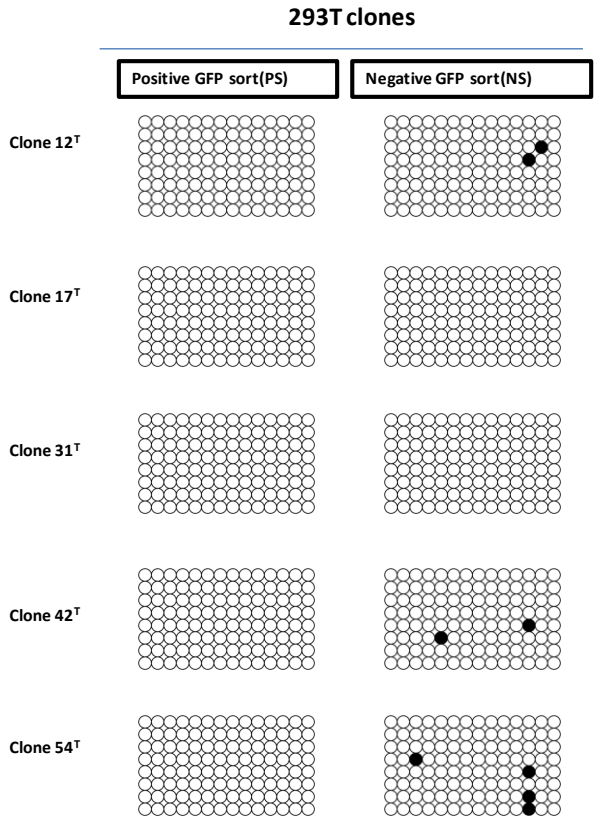
To analyze the DNA methylation status of the selected region of CMV promoter, bisulfite conversion of genomic DNA was done.Briefly, bisulfite treatment involves treatment of the DNA with sodium bisulfite which converts the cytosine into uracil. However, the cytosines that are methylated do not undergo this reaction and remain as cytosine. This converted DNA is that amplified by PCR using primers specific for the region of interest. PCR amplification leads to copying of uracils as thymine. The PCR products are cloned and then sequenced. The non-methylated cytosines are read as thymines whereas methylated cytosine remains cytosine.

Upon conversion of the DNA of the various cell populations, the CMV promoter fragment was amplified by PCR and integrated into a cloning vector and transformed in *E. coli*. From each HEK293T cell population eight bacterial clones were randomly picked and sequenced. Thus, the sequences reflect the methylation status of the CMV promoter in eight individual/single HEK293T cells.

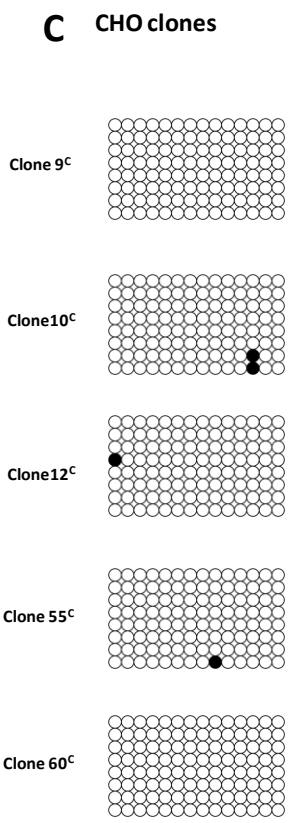
A


P1
Tgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggt
Ap1
gatgcgggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccc
CREB
attgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccc
Sp1
cattgacgcaaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctggtttagtgaac cgtc
P2
TSS
agatc

B


293T clones
Positive GFP sort(PS)
Negative GFP sort(NS)
Clone 12T
Clone 17T
Clone 31T
Clone 42T
Clone 54T

C


CHO clones
Clone 9C
Clone10C
Clone12C
Clone 55C
Clone 60C

D

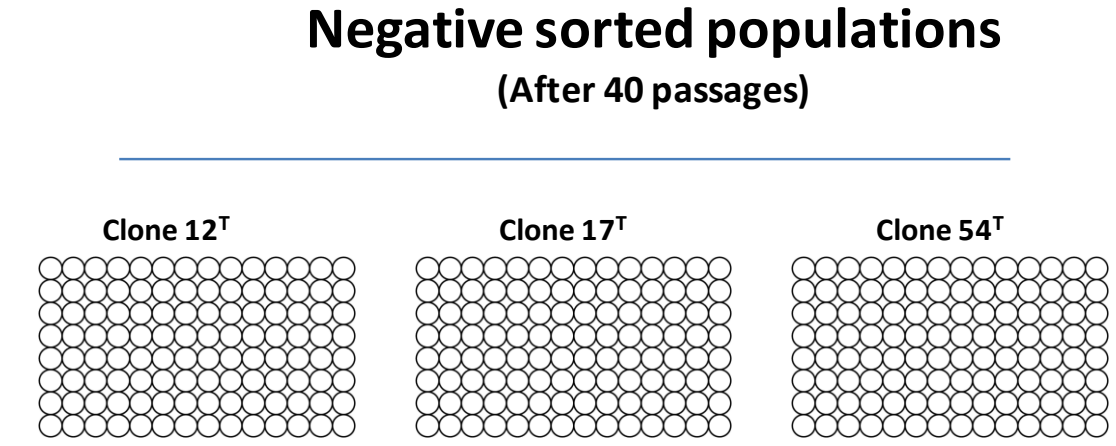

Negative sorted populations
(After 40 passages)
Clone 12T
Clone 17T
Clone 54T

Figure11: Methylation analysis of the CMV promoter. A) The sequence of the CMV promoter region used for methylation analysis. PCR primer binding sites are indicated by arrows (P1 and P2). The transcription factor and their binding sites are labelled and 14 CpG motifs are shown by dotted circle. The TATA box is highlighted in yellow. Genomic DNA was isolated from the sorted PS and NS populations of HEK293T (B) and 5 CHOK1 (C) clones and subjected to bisulfite conversion. The promoter region was amplified by PCR using the primers indicated in A and integrated into a cloning vector. Random clones were picked and sequenced. Each circle in a line represents a CpG dinucleotide according to the sequence above. The figure shows the sequences from 8 randomly selected clones per cell line reflecting independent single cells. Unfilled circle represent non methylated CpG dinucleotides and black filled circles represent methylated CpG dinucleotides..D) Analysis of NS population of clone12^{T}, 17^{T}, and 54^{T} after 40 passages.

None of the clonal population showed any methylated CpG in the GFP expressing cell fractions. This was as expected since promoter methylation is mostly associated with silencing and thus it should be absent in cells that express the transgene (Figure 11B).

However, contrary to expectations, sequencing of the clones from the cell populations that completely lost GFP expression (non-GFP expressing populations of clones 12^{T}, 17^{T} and 54^{T}; Figure 11B)revealed complete absence or rarely methylated CpGs (Figure 11B). Statistical analysis revealed that less than 0.5% of the analyzed CpG motifs in the CMV promoter fragment were found to be methylated. Thus, in none of the cell populations a significant number of methylated CpGs was observed.

To evaluate whether such a stable silencing without DNA methylation is a cell line dependent phenomenon or whether it is a more general mechanism adopted by the cells, analysis of representative clones derived from CHOK1 cells was also done. This included clone 12^{C} and clone 60^{C} (in which GFP expression rapidly dropped within the first two passages) and clones 9^{C}, 10^{C}, and 55^{C} with a heterogeneous expression pattern. Interestingly, in none of the CHOK1 clones methylation of CpGs played a major role (Figure 11C). Interestingly, also none of the clones from CHOK1 showed differential methylation pattern again contrary to what was expected since CHO 5 clones

To investigate if the methylation frequency would increase during cultivation the methylation status in these clones was also analysed after 40 passages post sorting. Importantly, also at this late time point both the expressing and non-expressing cells remained largely free of CMV promoter methylation (Figure 11D). This excludes a delayed manifestation of DNA methylation upon prolonged passaging as was previously suggested for other experimental settings (122-124).

Together, these results clearly show that silencing of gene expression in HEK293T cells and CHOK1 was not reflected by methylation of the CMV promoter. From these results it can be concluded that DNA methylation of the CMV promoter does neither cause nor correlate with the phenotypic heterogeneity and is thus not involved in silencing of CMV driven GFP expression in the HEK293T and CHOK1 derived cell clones.

3.1.5 High GFP expression correlates to differential levels of active and repressive histone markings

Since DNA methylation could be excluded to be the underlying mechanism of transgene silencing further analysis focused on the abundance of histone modifications in the sorted populations of the 5 selected clones. For this purpose chromatin immunoprecipitation (ChIP) was done. Briefly, in ChIP protein-DNA complexes are cross-linked, immunoprecipitated, purified, and amplified for gene- and promoter-specific analysis of known targets using real time PCR. The calculation of the enrichment is done through the 'percent input' method.With this method, signals obtained from the ChIP are divided by signals obtained from an input sample. This input sample represents the amount of chromatin used in the ChIP.

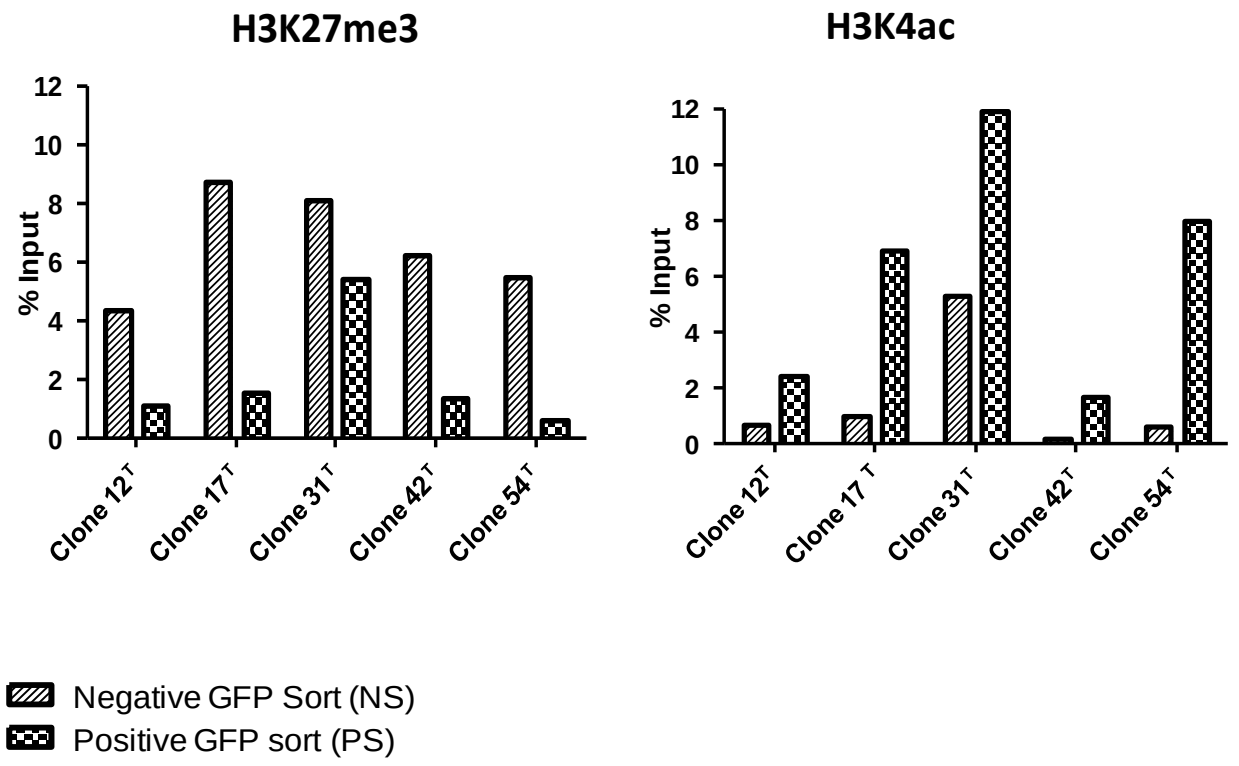


Figure12: Analysis of differential histone modifications in GFP expressing and non-expressing subpopulations. Chromatin immunoprecipitation (ChIP) was performed for the sorted PS and NS fractions from the five HEK293T clones using antibodies against H3K4 acetylation and Histone H3 lysine 27 trimethylation. The precipitated fractions were subjected to real time PCR. The quantification was done relative to percent input as described in the text and in the Material and Methods section. The graphs represent the mean from 3 separate ChIP experiments.

The acetylation of lysine 4 marking on histone H3 (H3K4ac) was evaluated using ChIP. H3K4ac has been described as a marker of open chromatin, which isassociated with transcriptionally active areas (125). Interestingly, the cell populations of the five clones displayed a differential pattern of histone modifications. The PS populations sorted for GFP expression were enriched for the H3K4ac marking. In contrast, the GFP negative NS population showed these H3K4ac marks with much less frequency (Figure 12). This indicates that GFP expression correlates with elevated levels of H3K4ac.

Since the differential pattern of acetylation was observed, another question was if there were other repressive differential histones marking also present with respect. To analyze this ChIP analysis was performed for trimethylation of lysine 27 on the histone H3 (H3K27me3) known to mediate DNA methylation. independent transcriptional suppression and is also held to be responsible for stabilizing the silencing phenotype (126). Following a ChIP analysis with a specific promoter according to the strategy described above, a strong enrichment of the H3K27me3 marking in the CMV promoter of the NS populations of all clones was observed. In contrast the PS population with GFP expressing cells showed lower frequency of this modification (Figure 12).

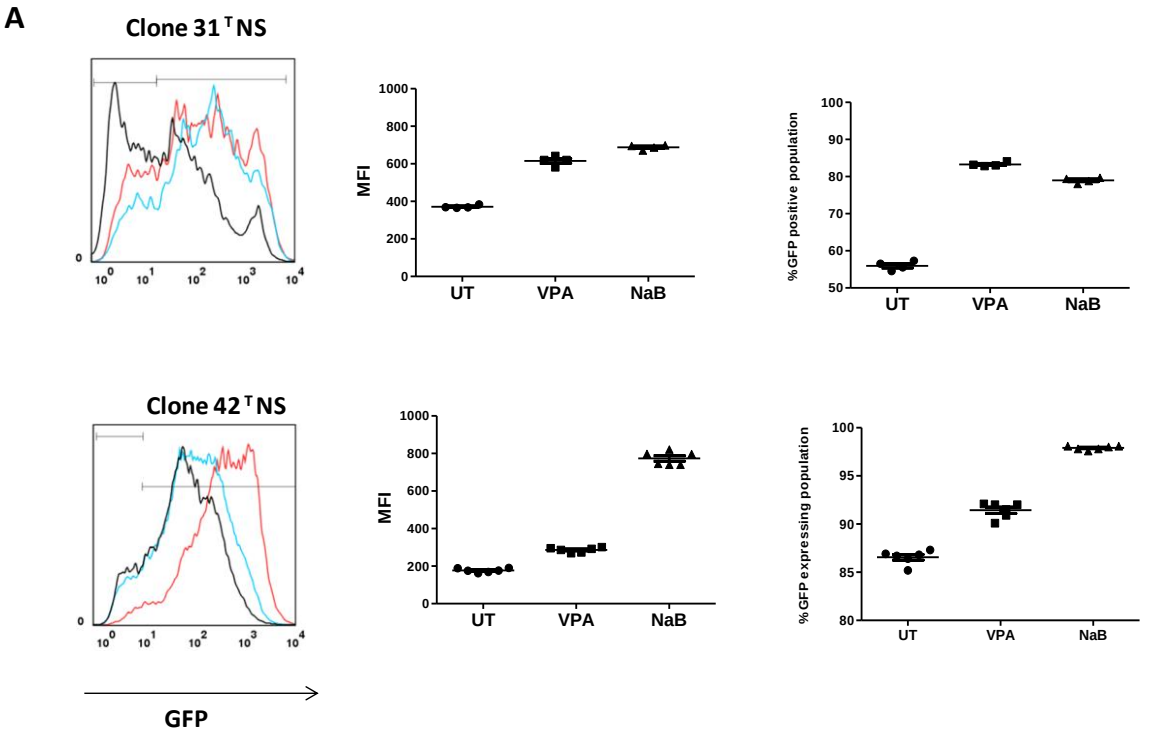
These results showed that the phenotypic expression patterns of the sorted populations correlate with a differential histone modification pattern.

3.1.6 Activation of transgene expression upon treatment with epigenetic modifiers

Since histone modifications were the major cause of silencing of the CMV promoter in the clonal populations, it was speculated that treatment of non-expressing cell populations with inhibitors of histone deacetylation (HDACi) is able to revert the silenced state and to increase the transgene expression. For this purposethe cells were cultivated for 72 hrs in presence of the HDACi sodium butyrate (NaB) (127) and Valproic acid (VPA) (128,129). Then, the expression level was evaluated.

Interestingly, the five clones responded differentially to the treatment with these drugs. In case of clone 31^T and 42^T these drugs induced a significant increase in the mean fluorescent intensity and in the percentage of expressing population (Figure13A).However, there was only a slight increase of the mean fluorescence intensity and the percentage of GFP expressing cells in the NS population of clone 54^T while the NS populations of clones 12^T and 17^T did not significantly revert to the expressing state (data not shown). This suggests that these populationswere locked in an epigenetically silenced state.

To confirm that the modification of the expression pattern upon treatment with these drugs wasreflected by changes in the histone marks of the CMV promoter we performed a ChIP assay for theNaB treated negative sorted populations of clones 31T and 42T. Incubation with NaB indeed enrichedthe chromatin with acetylation marking and caused a reduction in the H3 trimethylation levels (Figure 13B).



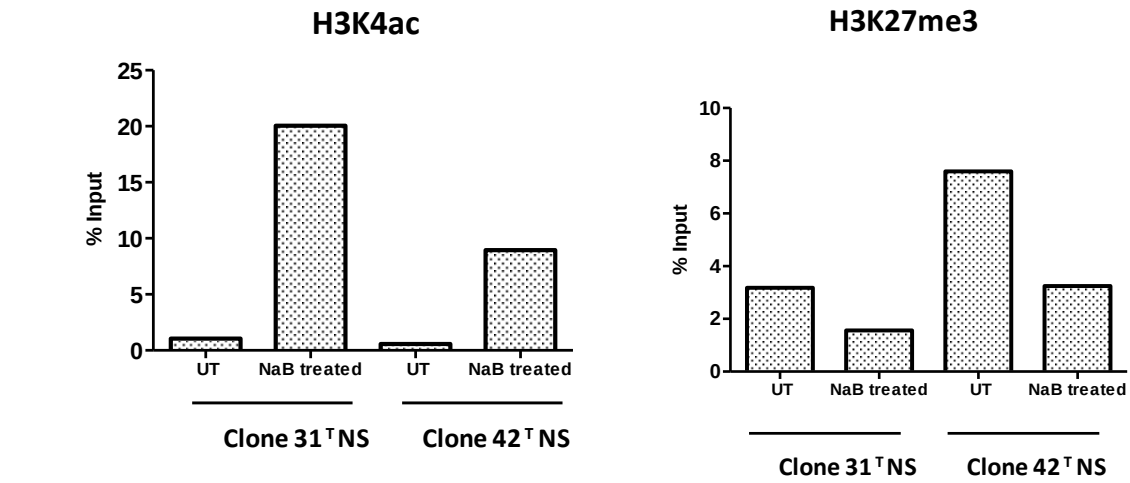


Figure13: Stability of histone mediated gene silencing. The indicated cell populations were treated for 72hrs with histone deacetylation inhibitors Valproic acid prior to analysis of GFP (VPA, blue line) and sodium butyrate (NaB, red line). UT indicates untreated controls (black lines). Overall GFP expression (left, representative populations), as well as % of GFP expressing cells (middle) and MFI (right) are given for 6 independent experiments. B) Treated cell populations were subjected to ChIP analysis for H3K4ac and H3K28me3. Graphs represent mean from 3 separate ChIP experiments related to % input

3.1.7 Reanalysis of the GFP negative and positive sorted populations

The different population of the five selected clones that were sorted in to GFP negative (NS) and GFP positive (PS) were cultured for 25 passages and were then reanalysed for their expression phenotype. This was done to analyze the stability of the two phenotypic variants from the clonal populations. Also it was be interesting to investigate if the differential histone marks can be responsible for stabilizing the silencing and maintain the silenced phenotype.

Interesting, the clones showed a different behaviour. The PS and NS subpopulations of clones 12^T, 17^T and 54^T showed a completely stable transgene expression phenotype: the GFP positive PS population remained positive and the GFP negative NS population also remained negative for GFP (Figure 14).

In contrast, both of the sorted populations of clones 31^T and 42^T revealed a highly heterogeneous expression pattern upon further cultivation: the NS populations of both clones shifted towards higher expression levels while the PS populations showed a partial loss of GFP expression. As a result, the respective populations merged upon passaging. This suggests that cells can undergo continuous changes from the non-expressing state to the expressing state and vice versa, thereby exhibiting a dynamic, “metastable” phenotype.

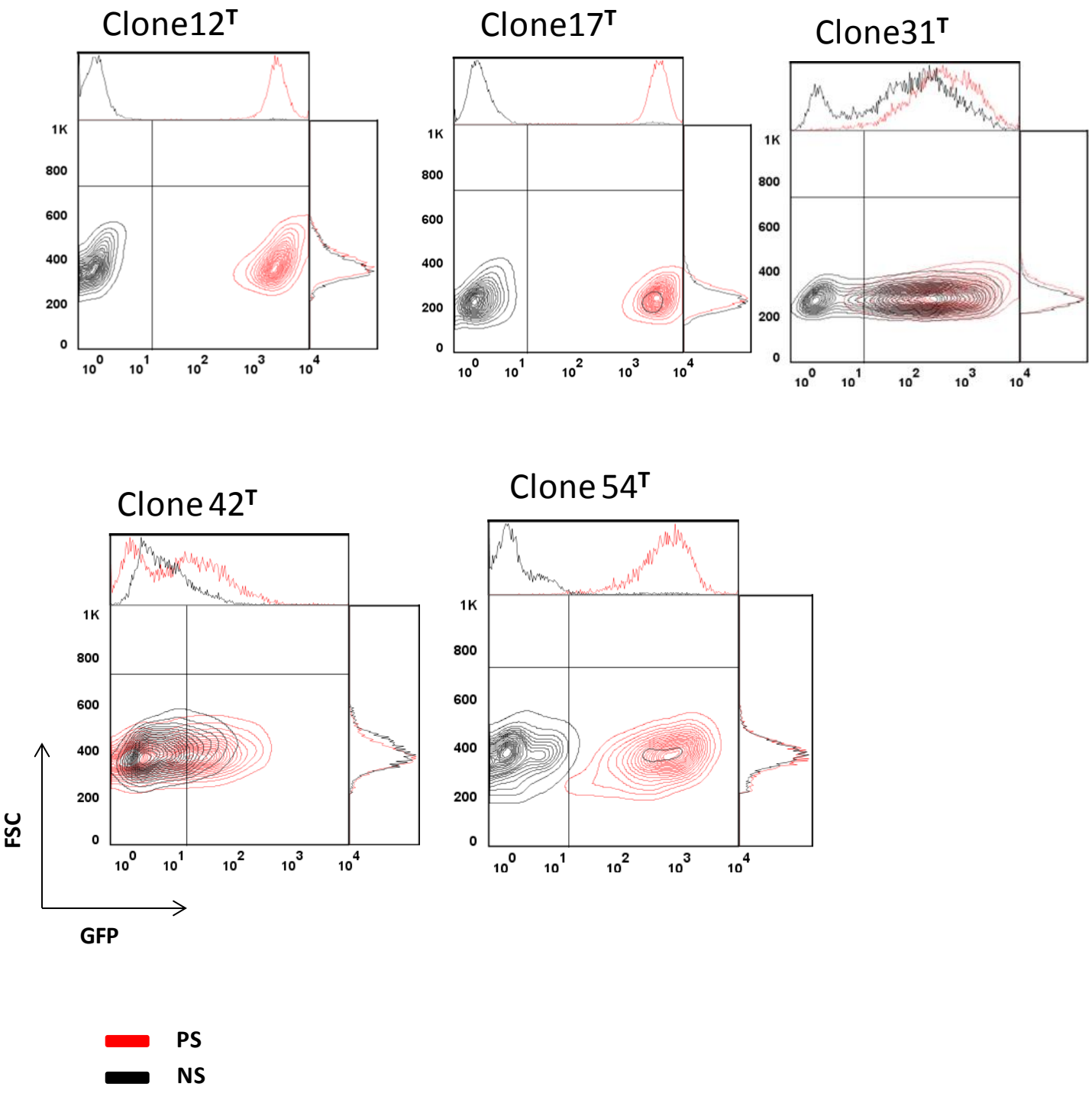


Figure 14: Stability of GFP expression in sorted clonal cell populations: 25 passages after sorting, the cell populations were subjected to expression analysis. The composite plots indicate the GFP expression of the PS and NS populations of the indicated cell clones

3.1.8 Expression profiles upon exchange of expression cassettes through recombinase mediated cassette exchange (RMCE)

Chromosomal hot spots of expression can be exploited by targeted integration of expression cassettes of choice (95,109,115,130-132). Thereby, it was possible to generate cell clones with predictable expression levels of a protein or even recombinant virus of interestCoroadinha, Schucht et al. 2006; Schucht, Coroadinha et al. 2006; Nehlsen, Schucht et al. 2009;.). So far, however, the chromatin status of the cassette upon exchange was not investigated. Thus, it was of interest to evaluate if the modulation of transgene expression at a particular site is a specific feature of the chromosomal locus and would be transferred also to incoming 'naked' DNA upon targeting. Evaluation was done for this potential in the selected GFP expressing cell populations, in particular the PS population of HEK293T clone 17^{T} as well as CHOK1 clones 9^{C} and 55^{C}. For this purpose the FRT-WT and FRT-F5 sequences were used for integrating a naive expression cassette by recombinase mediated cassette exchange (RMCE).

Three different cassette designs were analysed upon targeting. In the first one the coding gene GFP was replaced by RFP while the promoter was kept the same. In the second one a CMV based inducible Tet promoter was targeted. Finally, a constitutive, completely unrelated promoter (SV40) driving RFP was integrated. The results of these targetings are described in the following subchapters.

3.1.8a Exchange of the reporter coding gene

First, the cell populations were targeted with an FRT-WT/F5 flanked cassette comprising the CMV promoter driving RFP (CMV-RFP, Figure 15A). This would result in exchange of the reporter gene but maintenance of the same regulatory sequences. For this purpose the 1 x 10e5 cells of the respective sorted cell populations ($Cl17^{T}$, $Cl9^{C}$ and $Cl55^{C}$) were co-transfected with a plasmid encoding FlpE recombinase and the CMV-RFP reporter (Figure 15A). In particular, the FLPe expression vector pFlpe(116) was cotransfected with the targeting vectors pTAR CMV RFP (see section… (mat and meth) for details.In this system,Flp recombination can give rise to cassette exchange only since the CMV-GFP cassette in the isolated clones is flanked by a set of heterologous FRT sites and cannot recombine with each other (data not shown). Thus, excision of the GFP cassette is omitted. As a consequence of Flp mediated targeting of a CMV-RFP vector, targeted cells would lose GFP expression. Thus, excision of the GFP cassette is omitted. As a consequence of Flp mediated targeting of a CMV-RFP vector, targeted cells would lose GFP expression. Thus, 10 days after transfection, the cells were sorted for lack of GFP expression resulting from successful RMCE.

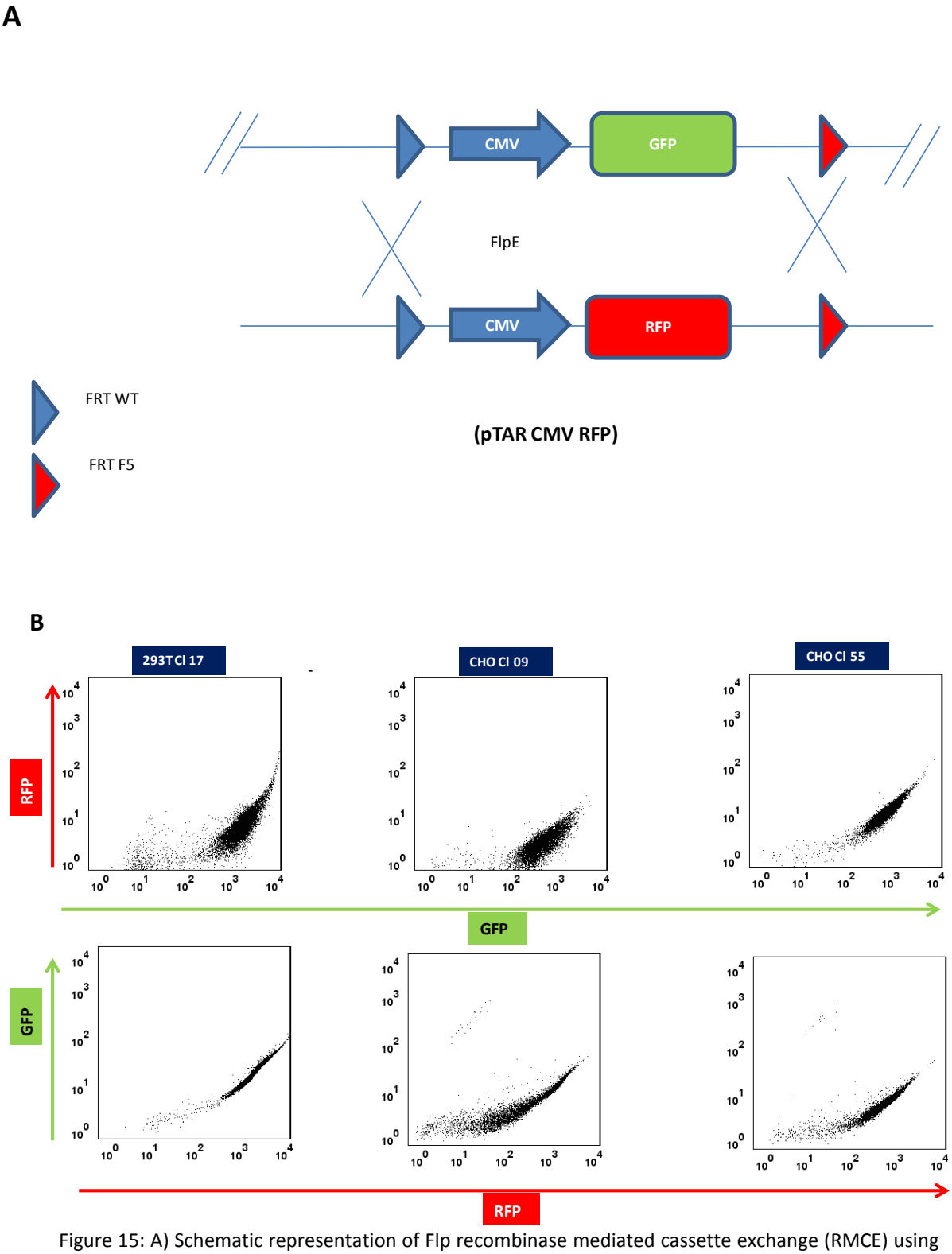


Figure 15: A) Schematic representation of Flp recombinase mediated cassette exchange (RMCE) using the vectors pTAR CMV RFP. B) Upon RMCE, for each clone, correctly targeted cells were isolated by sorting for loss of GFP expression. Sorted cells were subsequently evaluated for RFP expression by flow cytometry.The upper row of FACs plots in Figure 15B represents the parental cells prior to targeting and the lower row represents the sorted GFP negative cells after targeting

The sorted cell populations were evaluated for fluorescence, RFP after sorting. As expected, transfection of the recombinase plasmid pFlpe alone did not alter the percentage of GFP expressing cells (data not shown). When analysing the expression upon targeting of the clone 17^{T} PS, $Cl9^{C}$ and $Cl55^{C}$with the CMV-RFP vector, nearly all the sorted cells showed a high and homogeneous RFP expression, comparable to the GFP expression in the respective parental cells (Figure15B). These results show that upon integration of a naked DNA with different coding region but with the same promoter, the parental expression phenotype is restored. This might indicate to the possibility of that the major contribution in the crosstalk between the integration site and transgene might be played by the incoming promoter element in the transgene while the coding regions might not be playing that significant a role.

The mechanism of such maintenance of the parental phenotype might lie in the successful inheritance of the favorable epigenetic signatures that were present in the parental populations and might be the reason of presence of the memory phenotype a observed in the above cases.

3.1.8b Evaluation of a synthetic promoter construct in HEKCl17PS site

The previous chapter indicated that the clones gave rise to reproducible expression when a transgene cassette was introduced with different coding gene but with the same promoter. The interesting observation that came out from the above experimental settings was that the integration site supported expression of different genes from CMV promoter and these integration sites were favorable for transgene expression from CMV promoter in a constitutive manner. The CMV promoter has also been used to create the tetracycline inducible synthetic inducible Tet promoter (80). Thus, it was evaluated if these chromosomal sites are not only supporting constitutive expression from the CMV promoter but are also good of inducible expression from CMV derived tetracycline promoter. To investigate if the screened sites might also be supporting high, and strictly regulated expression from a Tet inducible system based on the minimal CMV promoter, a synthetic cassette, pEM-rtT2luc3eGFP(99)having a bidirectional Tet promoter driving GFP and luciferase (Figure 16A), was used. To facilitate screening for cassette exchange, this vector was targeted into the Cl17PS CMV RFP cells (discussed in Figure15B).

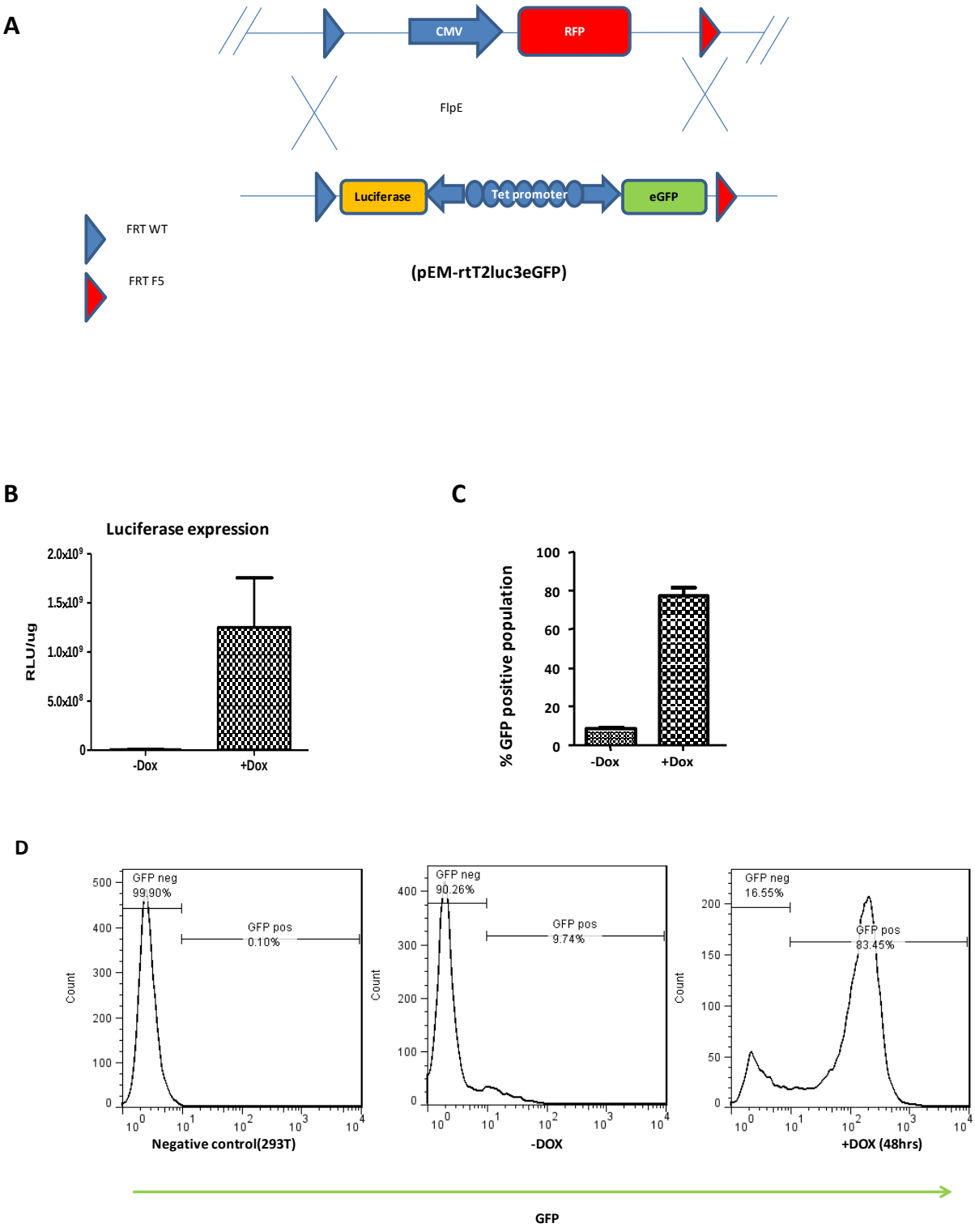


Figure 16: A) Schematic representation of Flp recombinase mediated cassette exchange (RMCE) using the vector pEM-rtT2luc3eGFP. BUpon RMCE, correctly targeted cells were isolated by sorting for loss of RFP expression since the BiTet had GFP as one of the markers so the CL17 with CMV RFP was targeted. Sorted cells were evaluated for luciferase expression in the on state and (+Dox) and in an un-induced state (-Dox). C) The quantification of GFP expression in on and off state as measured by flow cytometry analysis. (Error bars represent the deviation in three independent experiments). D) FACS plots showing heterogeneous expression from the BiTet promoter after even 48 hr of induction.

Correctly targeted cells were identified and sorted for lack of RFP expression. The generated cell populations were then evaluated for luciferase expression upon treatment with Dox. For this, cells were kept in presence of Dox (2mg/ml) and as a control without Dox in the culture medium. Cell lysates were analyzed for luciferase expression. Also the expression of the GFP was also measured through flow cytometry. While the luciferase expression is highly sensitive, the GFP expression can be analysed through flow cytometry giving precise single cell expression analysis.

Interestingly, the integration site from Cl 17 PS did support inducible expression from the bidirectional Tet promoter after 48 hr of Dox treatment. High luciferase expression ($>10^9$RLu/µg) (Figure16B) and inducible GFP expression with nearly 80% of cell undergoing induction upon Dox treatment (Figure16C) was observed. This was highly interesting and contrary to expectation since not only is Tet promoter susceptible to silencing but also if targeted in high expressing sites, the regulated expression is generally lost(data not shown).However another interesting observation was that even after induction for 48 hr there were still about 20 % cells in the population that were not induced. This showed that not only in the constitutive CMV promoter, heterogeneity was also evident for CMV derived Tet promoter driven construct.

Therefore, this integration site was not only able to support constitutive expression but was also good for inducible expression with induction of nearly 80 % of the total cell population. However, the cell to cell heterogeneity still existed even in the transgenes driven by BiTet promoter as observed upon single cell analysis through flow cytometry.

3.1.8cTargeting a heterologous construct in the tagged loci in HEK293T and CHOK1 cell

The question than was asked what kind of expression phenotype would be generated upon integrating a heterologous promoter into the same chromosomal sites. To answer this, targeting was done with a SV40-RFP cassette in which the SV40 promoter drives RFP (pTAR SV40 RFP) (Figure 17A).

Correctly targeted cells were identified and sorted for lack of GFP expression according to the experimental protocol outlined in chapter 3.1.9. Interestingly, different outcome were observed in the three cell populations targeted by the SV40-RFP cassette. While for the CHOK1 clone 9^{C} the high level of expression of the parental population was maintained, a highly heterogeneous expression was observed in clonal populations of clone 17^{T} and clone 55^{C} (Figure 17B). Both in clone 55^{C} and the PS population of clone 17^{T}, a significant fraction of cells could not express RFP although they were successfully targeted.

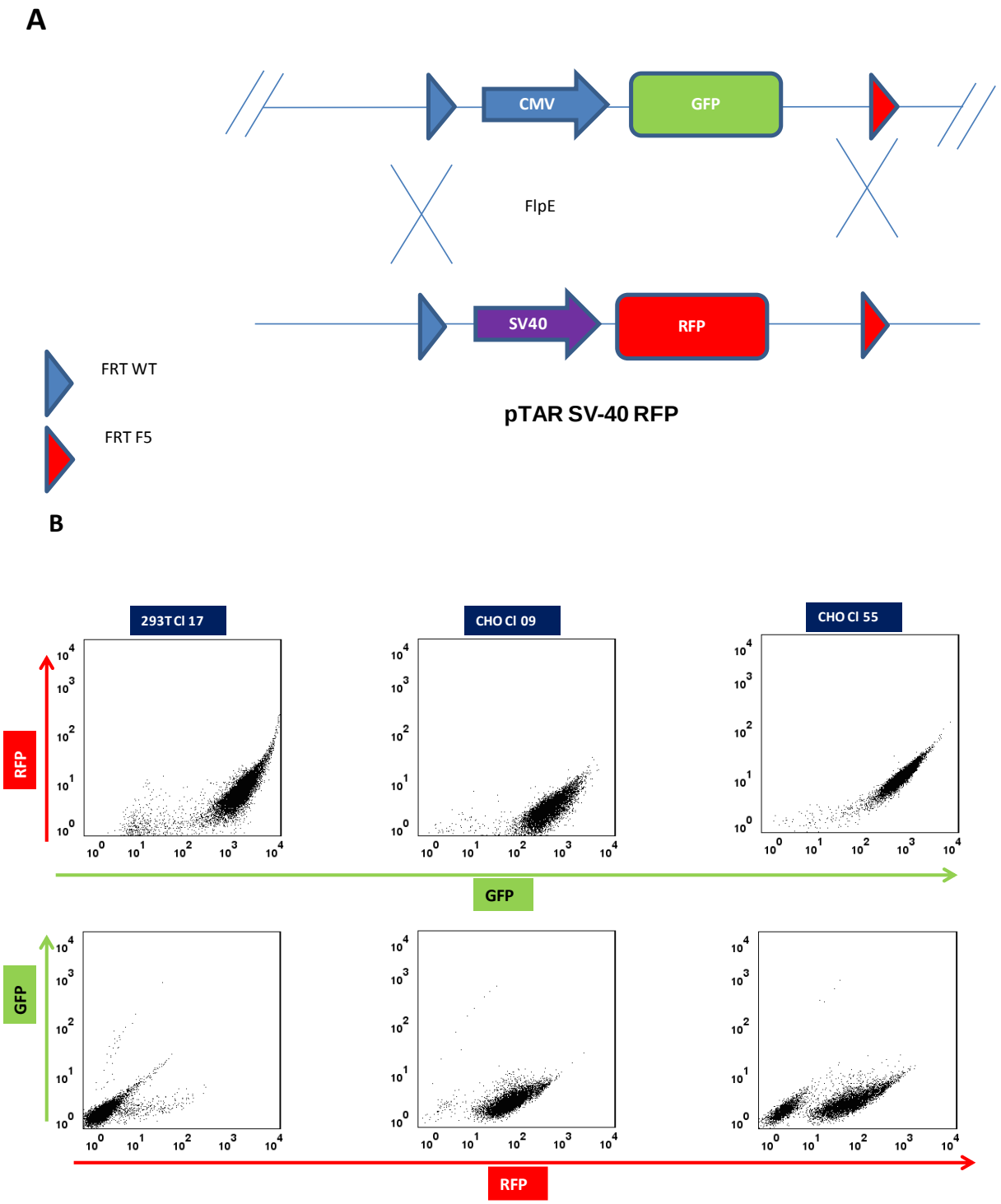
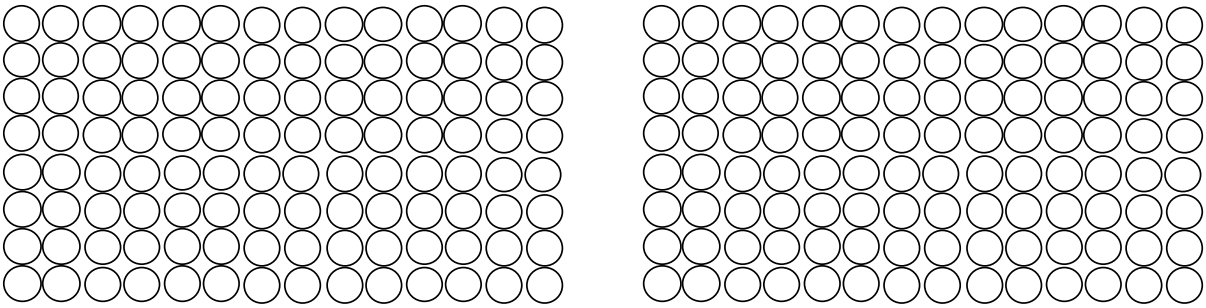


Figure17: A) Schematic representation of Flp recombinase mediated cassette exchange (RMCE) using the vectors pTAR SV-40 RFP. B) The clones 17^{T}, 9^{C} and 55^{C} were used. Upon RMCE, for each clone, correctly targeted cells were isolated by sorting for loss of GFP expression. Sorted cells were evaluated for RFP expression by flow cytometry.The upper row of FACs plots in Figure 17B represents the parental state prior to targeting and the lower row represents the state after targeting

3.1.8d Evaluation of the methylation state of the SV40 promoter upon targeting in high expressing sites

The introduction of naked DNA elements in the previous chapters showed different expression patterns which suggested that the crosstalk of the expression cassettes with the chromosomal environment plays a crucial role in determining the expression from a transgene introduced in the genome. However, the question was if the underlying epigenetic mechanisms were same or different. In particular, this concerned the situation/the case where the introduction of the heterologous promoter lead to silencing or heterogeneous expression,. In chapter… it was demonstrated that in the parental these sites as such were free from the effect of DNA methylation but showed differential histone marks (see Figures 11 and 12 for Cl17 NS and PS, respectively). Thus, it was of interest to investigate if methylation sets in after targeting of the SV40 promoter cassette and contributes to the partial loss of expression.

Thus, the non-expressing and expressing cell fraction of SV40 RFP targeted clone 55^{C} and clone 17^{T} cells were sorted and subjected to DNA methylation analysis of the SV40 promoter according to the protocol outlined in chapter 3.1.4. Interestingly, from the 14 CpGs evaluated, none of them was affected by DNA methylation (Figure 18).



Cl17^{T} **Cl55^{C}**

Figure 18: DNA methylation analysis of silenced fraction of SV40 driven RFP targeted populations from $Cl17^{T}$ and Clone 55^{C}.Genomic DNA was isolated from the negative sorted populations of HEK293T (Cl17) and CHOK1 (Cl55) clones and subjected to bisulfite conversion. The promoter region was amplified by PCR and integrated into a cloning vector. Random clones were picked and sequenced. 14 CpG motifs from the SV40 promoter fragment are shown by circles. Each circle in a line represents a CpG dinucleotide. The figure shows the sequences from randomly selected clones reflecting independent single cells. Unfilled circle represent non methylated CpG dinucleotides and black filled circles represent methylated CpG dinucleotides

3.1.9 Summary of results on targeted integration

The evaluation of the expression upon integration of heterologous promoter SV40, interestingly, revealed a very interesting expression profile. Staring from the complete loss of the complete expression in 293T Cl17 PS suggested that this site is not good for SV40 driven transgene expression. However, other site such as in CHO $Cl9^{C}$ showed maintenance of parental expression phenotype. Such kinds of sites are rare but are important since these might be the sites that can favour transgene expression from different transgenes. CHO $CL55^{C}$ represented an intermediate between the two extremes where the integration resulted into the clone becoming unstable with a distinct population having lost expression transgene, however it still had a population that expressed the transgene.

Together, the three clones investigated here nicely represent various expression patterns namely stable maintenance of expression, stable silencing and heterogeneous expression. All the three sites showed good expression of CMV driven transgenes with high and stable expression of both GFP and RFP (Figure 15). This not only justified the strategy used for screening (lentiviral transduction and sorting for high expression on day10) but also suggested that the CMV driven expression might also be used to screen for sites that can support strict inducible transgene expression as shown for Cl17 PS (Figure16). This is helpful since the inducible expression from the CMV driven Tet promoter can have stochastic activation and this might pose problems in screening for good integration sites.

Another interesting observation that came out is the prospect of locating a site that can be used to universally express different transgenes. The Cl9C represented one of the few sites that showed uniform and stable transgene expression from both types of promoter.

Although, the promoter strength might vary in a particular site but the uniform expression was highly promising. The DNA methylation independent silencing of transgene expression was quite interesting. On one hand, this suggested that the SV40 can also be silenced in DNA methylation independent manner while on the other hand it also suggested that these sight are more prone to getting silenced by mechanism that donot involve methylation of DNA.

Overall, this section of the thesis evaluated different integration sites and the transgene expression. Most of integrations sites do not have homogeneous expression and are marred by some levels of heterogeneity. However, once screened for high expression using a certain promoter-transgene combination might result into screening for integration sites that are specific for this combination. This is also reflected by the epigenetic pattern that is imbibed by the transgene and similar epigenetic might be inherited by other transgenes as long as the promoter –integration site are kept the same.

However, upon integration of a different promoter in the screened site might no longer be fit for stable expression. The expression of transgene from new promoter might be dictated by the new kind of crosstalk.

3.2 Results Part 2

3.2 Crosstalk between synthetic construct in the Rosa26 locus

The previous chapter described the modulation of the well-known and widely used CMV promoter in different integration sites in biotechnologically relevant CHO and HEK293T cell lines. A similar situation arises upon genetic modification of stem cells when generating transgenic animal models. In this chapter, the focus will be on the crosstalk arising from the integration of a Tetracycline inducible promoter (Tet promoter) in the well-characterized Rosa26 locus of mice (Soriano et al,). In particular, this focus is laid on the evaluation of the RosaGFP mice.

3.2.1Expression analysis of RosaGFP mice

TheRosaGFP transgenic mouse line was previously generated upon integrating a synthetic bidirectional Tet (BiTet) promoter cassette into the Rosa26 locus by RMCE (99). For this purpose, embryonic stem (ES) cells were used in which the Rosa locus had been modified by the heterospecific FRT sites (FRT-WT and FRT-F5)(see Figure 19A). RMCE was used to integrate a cassette comprising the BiTet promoter driving GFP and luciferase. This cassette also includes the gene for the BiTet transactivation (rtTA) which is expressed from the ubiquitously active Rosa 26 promoter upon RMCE (Figure 19B). Thus, this cassette comprises all elements required for Dox controlled expression in all tissues in the mouse.Therefore in this Rosa GFP transgenic mouse model, administration of Doxycycline results in binding of the rtTA to the BiTet promoter, thereby leading to activation of transcription from the BiTet promoter. Hence, upon Doxycycline administration ubiquitous expression of GFP and luciferase would take place. As a control, the theConL mouse line was used. In this line, the luciferase gene is integrated into the Rosa26 locus (Sandhu et al., 2011).

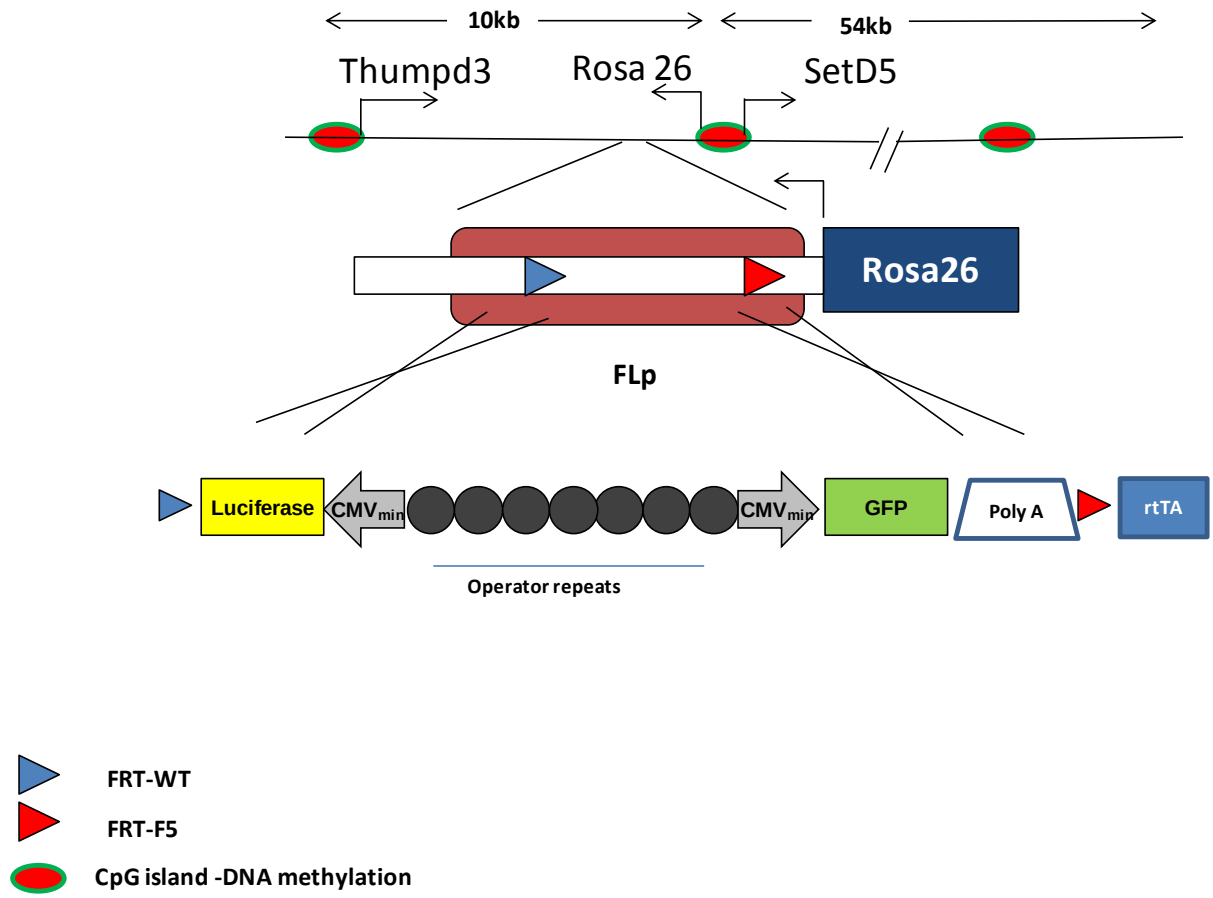


Figure 19: Schematic depiction of the genomic structure of the Rosa26 locus (not to scale) and the neighbouring endogenous promoters along with the CpG islands. This locus was made RMCE competent by implementation of FRT-WT/FRT-F5 sites as indicated (99)On the level of ES cells, a synthetic cassette has been integrated into this locus.The cassette comprises of a synthetic Tet promoter (BiTet) driving luciferase and GFP, and rtTA gene which would driven by the Rosa26 promoter upon targeting). From these ES cells, a transgenic mouse line was generated, designated as RosaGFP (Sandhu and Wirth, unpublished)

To evaluate gene expression, Rosa-GFP mice were initially kept without Doxycycline and were measured via in vivo bioluminescence (see chapter 2.2.10 for the details). Only residual basal levels of luciferase expression were determined. This ranged below 2000 p/s/cm2/sr.To switch the BiTet promoter into transcriptionally active form and induce the expression of GFP and luciferase, these mice were then fed with 0.2mg/ml of Doxycycline in the drinking water for more than 2 weeks (18 days). These mice were then monitored for the luciferase expression via bioluminescence using *in vivo* IVIS imaging. As a comparison, RosaConL mice with constitutive expression of luciferase from the Rosa26 promoter were included.

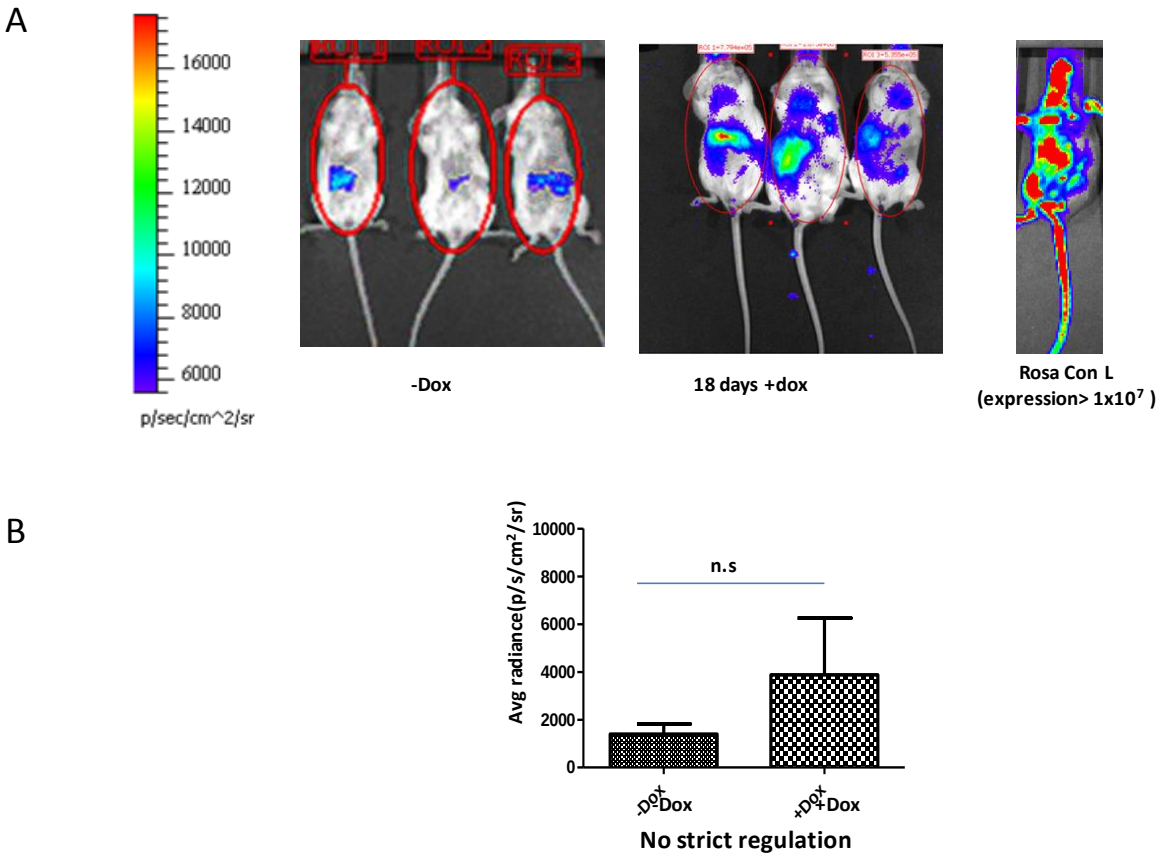


Figure 20: Luciferase expression inRosaGFP mice: A) BiTet driven expression of luciferase in Rosa-GFP mice was quantified using *in-vivo* IVIS imaging. The expression of luciferase was quantified in mice before and after feeding with 0.2 mg/ml Dox in drinking water for 18 days. The control for luciferase was constitutively luciferase expressing ConL mice with luciferase under the Rosa26 promoter (99). B) The graph shows the luciferasequantification in the two mice groups i) without Dox ii) with Dox (n=6, significance was calculated usingMann Whitney test, p=0,0649)

Surprisingly, in the induced state the bioluminescence signal was only slightly induced from 1400(p/s/cm^2/sr) to 3879 (p/s/cm^2/sr). In contrast, in the Con L mouse a signal of 10E7-10E8 (p/s/cm^2/sr) was detected (see representative pictures in Figure 20A).

When comparing the expression in individual mice, some amount of mouse to mouse variation was observed (Figure 20A and data not shown). However, in none of the mice pronounced induction of luciferase expression was observed.Statistical analysis revealed that the mice did not show a significant regulation of the luciferase expression in the BiTet on(+Dox) and BiTet off (-Dox) state (Figure 20B).

3.2.2 Partial rescue of the transgene expression upon treatment with Dnmt inhibitors

The lack of expression of the luciferase as measured from bioluminescence in the induced RosaGFP transgenic mouse model was in sharp contrast to the high expression of control RosaConL mice which had constitutive luciferase expression from the Rosa26 promoter. It was hypothesisedthat epigenetic influences could lead to the lack of transgene expression in the Rosa26 site. Epigenetic modulation could affect either the BiTet promoter or the Rosa26 promoter driving the rtTA transactivator.

In a first analysis it was evaluated if transgene expression is affected by DNA methylation. Therefore, the role of DNA methylation in the silencing of Tet driven transgene in the Rosa locus was first to be investigated. For this purpose, in this study Azacytidine (Aza) and Decitabine (Deci) were used since they specifically impair DNA methyltransferase (Dnmt) activity. In order to investigate the possible role of DNA methylation in lack of Tet driven transgene expression, the RosaGFPmice were investigated upon treatment with the DNMT inhibitors Aza and Deci (0.5mg/ml each, intraperitonally) for two consecutive days in presence and absence of Doxycycline (groups 3 and 4 in Figure21). As a control, mice without DNMT treatment were analyzed in presence and absence of Doxycycline (groups 1 and 2 in Figure 20). Since demethylation is restricted to actively proliferating cells, Rosa-GFP mice from group 3 and group 4 were additionally treated with 10 µl/gcarbon tetrachloride (CCl_4) at day 0. CCl4is known to cause liver damage after being metabolized (133), thereby inducing proliferation in the liver cells. The overall scheme is summarized in Figure 21A.

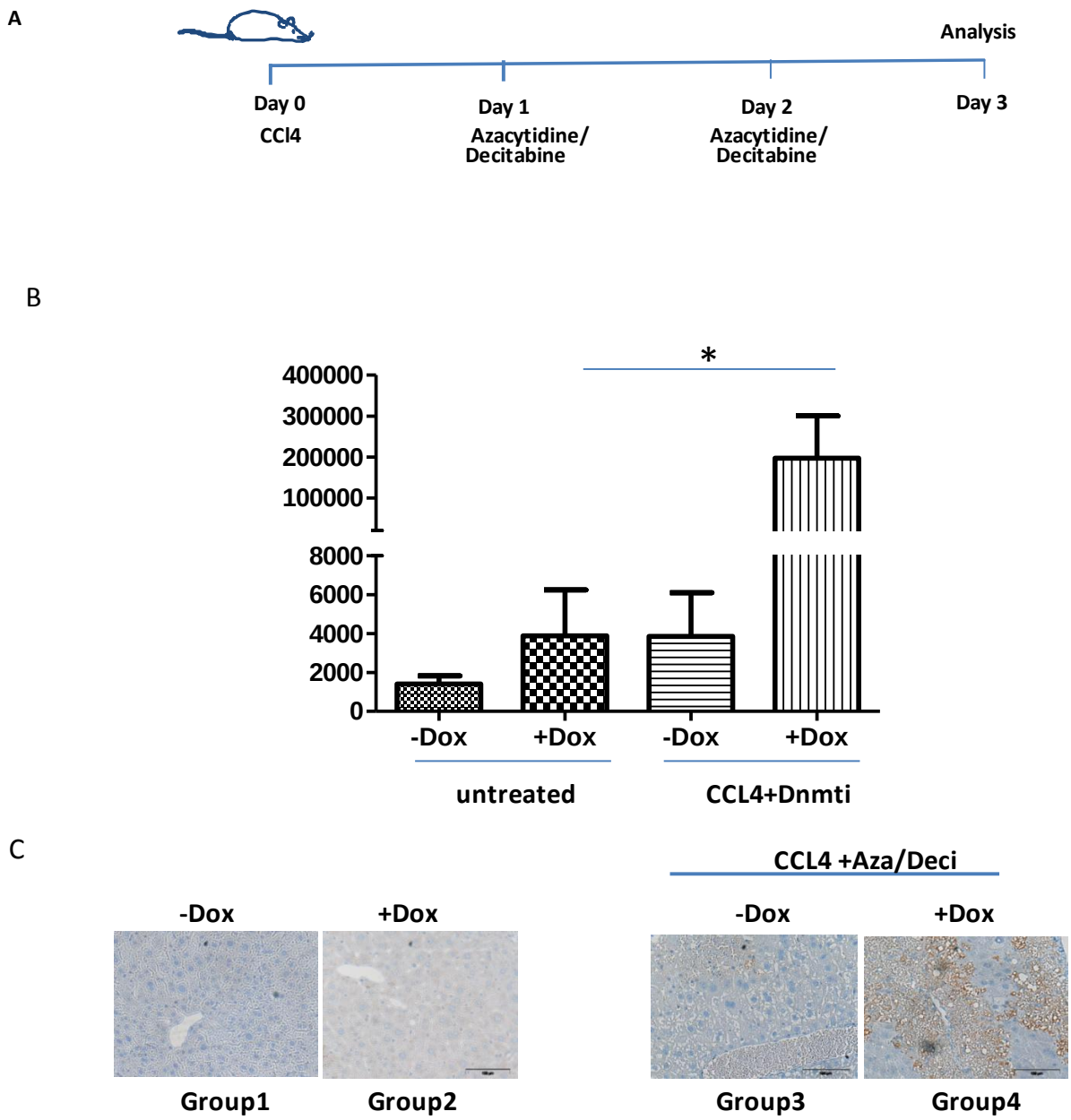


Figure 21: Transgene expression pattern of RosaGFP mice upon treatment with Dnmt inhibiting drugs: A) Depiction of experimental procedure with mice being subjected to CCl_4(10µl/g) induced liver damage on day 0 and consecutive treatments with Azacytidine and Decitabine (0.5mg/ml each) on day 1 and day 2 followed by IVIS measurements on day 3. B) The graph shows the average radiance of non-induced mice (-Dox), induced mice (+Dox) and uninduced mice with additional treatment of Dnmt inhibitors upon CCl4 mediated liver damage(-Dox+CCL4+Azacytidine+Decitabine) and induced mice with additional treatment of Dnmt inhibitors upon CCl4 mediated liver damage (+Dox+CCL4+Azacytidine+Decitabine) (p=0.02 using Mann Whitney Tet). C) Mice were sacrificed and characterized by histology staining of liver tissue sections using an anti-GFP antibody. Representative pictures from mice of all the four groups are shown. Scale bar100um, each group had 4 mice (n=4)

To confirm the induction of proliferation upon CCL_4 mediated liver damage, the sections were stained for Ki-67 expression. Ki-67 is a proliferation marker. For this purpose tissue sections from liver, spleen and kidney of the 4 mouse groups were used. Indeed, only the liver sections of CCL_4 treated mice (group3 and group4) showed profound staining for Ki-67 (data not shown) indicating active proliferating liver tissue. No significant staining was observed in the other organs (data not shown) which was expected since liver is the primary site of injury post CCL_4 treatment.

The luciferase expression was measured on day 3 (Figure 21B) after the initial CCL4 treatment via bioluminescence using *in vivo* IVIS imaging. A significantly increased luciferase expression by nearly 100 fold was observed in group 4 (+Dox/+CCL_4/+Aza +Deci) in comparison to untreated control group mice group 2 (+Dox only) (Figure 21B).

Further to evaluate the expression of GFP, the animals were sacrificed and tissue slides from liver were stained with an anti-GFP antibody. No expression was observed in groups 1, 2 and 3 as well as in control animals treated with CCl4 only (Figure 21C and data not shown). However, samples from group 4 showed broad staining of hepatocytes with large number of hepatocytes showing positive GFP staining. This indicated that the expression seen after the treatment was specific and strongly suggested the prevalence of DNA methylation in the Tet driven transgene expression in the Rosa 26 locus. Interestingly, in these conditions not all hepatocytes showed transgene expression

Thus, a partial rescue not only in luciferase expression but also in GFP expression upon treatment with azacytidine and Decitabine in mice from group 4 was observed. This observation strongly supported the hypothesis of DNA methylation mediated silencing of the Tet driven transgene might be occurring in the Rosa locus. Further chapter focused on analyzing whether the Tet promoter was silenced leading to loss of transgene expression or the Rosa26 promoter was silenced, leading to loss of transactivator expression.

3.2.3 DNA methylation pattern of the BiTet promoter in the Rosa26 locus

The RosaGFP mice showed hardly any expression in all organs (Figure 20 and data not shown). The treatmentwith Azacytidine and Decitabine resulted in partial rescue of GFP and luciferase expression in Rosa-GFP transgenic mice. This strongly suggested the possibility of BiTet promoter and Rosa26 promoter being silenced due to methylation of it CpG motifs.

Firstly, the BiTet promoter was analyzed. It has a high density of CpGs with 5 CpGs each in the minimal CMV promoter regions and one CpG in each of the linker DNA connecting the seven tetracycline operator repeats, together resulting in 17 CpGs (Figure22A).

To analyse the methylation status of the Tet promoter DNA samples from liver were taken from three individual RosaGFP mice and were subjected to bisulfite treatments per the protocol described in chapter3.2.2. A specific primer set was used to amplify the whole BiTet promoter. PCR fragments were cloned and a minimum of 8 random clones were analysed. After sequencing, the methylated cytosines in the BiTet promoter were determined.

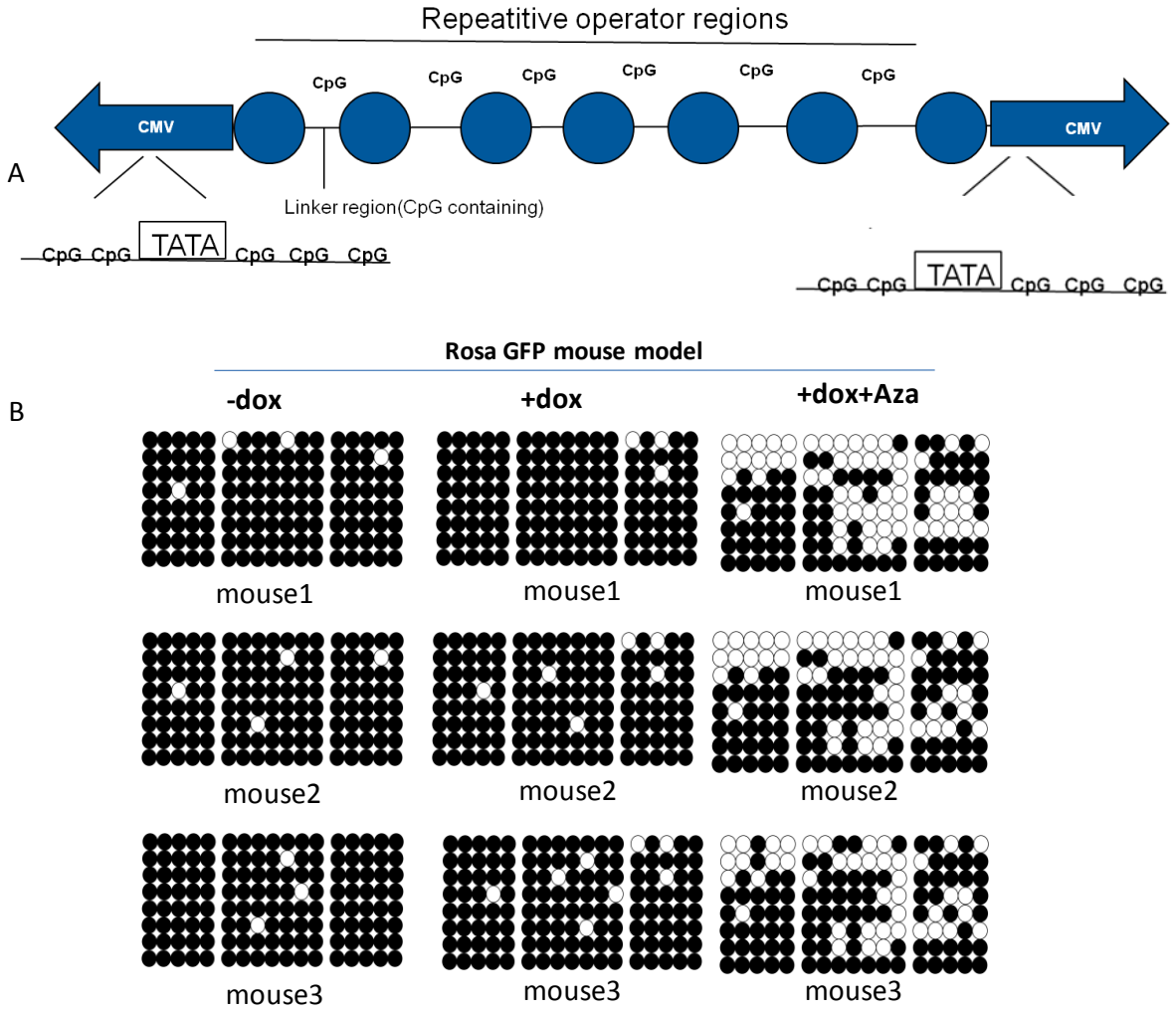


Figure 22: DNA methylation analysis of the BiTet promoter. A: depiction of the BiTet promoter. The position of the 17 CpG motifs is indicated. B: Genomic DNA was isolated from the liver tissues of three Rosa-GFP mice and subjected to the indicated treatments (see Figure 21 for description of the treatment). After bisulfite conversion the promoter region was amplified by PCR. The amplified product was integrated into a cloning vector. Random clones were picked and sequenced. The methylation pattern of three mice per group is depicted. Each circle in a line represents a CpG dinucleotide, the complete row of 17 circle represents the BiTet promoter in a single cell. Eight promoter fragments per sample are depicted. Unfilled circle represent non methylated CpG dinucleotides and black filled circles represent methylated CpG dinucleotides

The methylation analysis revealed high levels of methylation both in untreated and Doxycycline treated animals (group 1 and group 2 in Figure 20). The three group 1 animals showed methylation frequencies of 93-97%, the Doxycycline induced group displayed 91-94% methylation. In contrast, the mice subjected to CCL_4 and Dnmti treatment overall methylation was found to be indeed decreased significantly with 63%, 68% and 65% for the three animals.

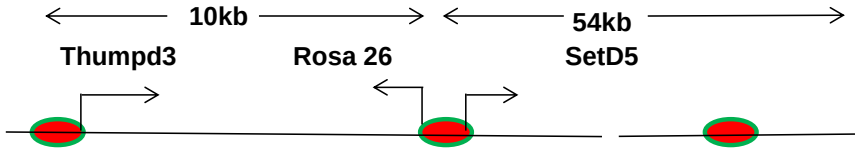
Importantly, the treatment with CCl_4 + Aza did indeed result in the partial demethylation of the BiTet promoter. This suggests that indeed epigenetic silencing is on the basis of the BiTet promoter. The results nicely correlate with the observed rescue of luciferase expression and the GFP expression (cf. Figure 22). Still, a large number of CpGs remained methylated which corresponds to a significant number of non-expressing cells as revealed by the histology (Figure 21C). This could be the consequence of limited efficiency of such treatments. Alternatively, other factors might also play a role in bringing about silencing of BiTet.

3.2.4 Endogenous Rosa26 promoter and Thumpd3 promoter remained methylation free upon targeting the Tet driven constructs.

Theoretically, lack of expression of the BiTet promoter might have been a consequence of silencing of the Rosa26 driven rtTA gene that is required for transcriptional activation of the BiTet promoter.The locus harbours two classical CpG island promoters in close proximity, namely Rosa26 and Thumpd3 promoter (Figure 23A). The endogenous Rosa26 locus is known to show a ubiquitous expression pattern in various developmental stages. High levels of polymerase II occupancy and high levels of H3 acetylation were observed for the Rosa26 promoter upon chromatin immune precipitation (ChIP) in non-modified mouse NIH3T3 cells (data not shown). However, different studies suggested the possibility of spreading of the DNA methylation which could affect also such an endogenous promoter (134-136): according to this study highly methylated sequence can trigger the spread of methylation in the neighbouring sequence. Thus, the question arose if thehighly methylatedBiTet promoter could trigger the methylation of the endogenous promoters that flank the transgene integration site.

To answer this question, DNA methylation analysis of the endogenous Rosa26 locus was performed. Thus, bisulfite treatment of DNA isolated from liver, lung, kidney and spleen of RosaGFP mice wasperformed. In this experiment, another method was applied to assess the level of methylation. Instead of cloning the PCR amplified fragments, a mass spectrometric analysis was performed using Epitypher (Sequenom). In contrast to the analysis of few PCR fragments representing individual cells the Epitypher analysis provides quantitative data relying on a large number of reads/data points per sample allowing statistical evaluation.

A



B

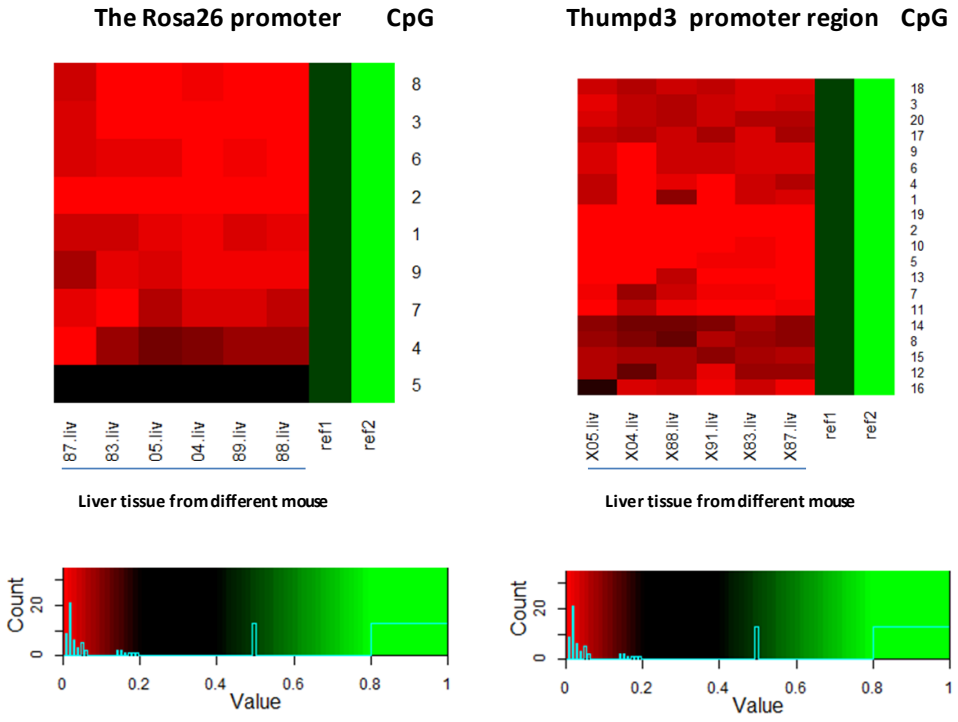


Figure 23: DNA methylation analysis of the Rosa26 locus. A) Map of the Rosa26 locus harbouring two classical CpG island promoters; the Rosa26 promoter and the Thumpd3 promoter. B) The heat maps represent the results of the methylation analysis done on Epitypher. The reference scale is shown below. Red represents 0-15%, dark green 50% (0.5) methylation and light green being 100% (1.0) of methylation. The various CpG motifs of the two promoters are indicated on the right side of each map. Ref1 represents the reference for in vitro methylated fragment with 50% methylation while ref2 is the 100 percent in vitro methylation reference. Each column represents different CpG from a single liver sample from independent mice while each row represents same CpG analysed for different samples.

In contrast to the fully methylated control (Ref 2 in Figure 23), the quantitative DNA methylation analysis from liver revealed very low levels of overall methylation in both the endogenous promoter regions Rosa26 and Thumpd3 (Figure 23B). The results of the analysis from other organs such as spleens, lungs, kidney were comparable (data not shown). This indicates that the endogenous promoters remain methylation free even if they are flanked by a highly methylated BiTet promoter element. Together, these results clearly indicate that

the spreading phenomenon as suggested by certain studies (134-136) does not occur in the Rosa locus.

Together, with the results obtained in the previous chapter 3.2.4; this suggests that the high levels of methylation are restricted to the transgene with hyper methylated Tet promoter not having any affect on the expression of the Rosa26 driven rtTA gene.

3.2.5 Partial rescue of GFP expression upon treatment with Azacytidine and Decitabine in the RosaGFP mES cells

In vivo based expression analysis and also the ex vivo histological staining confirmed the role of Dnmti in rescuing the transgene expressing. Further the DNA methylation analysis of the liver tissues showed high levels of methylation in the BiTet promoter. The partially demethylated state obtained upon treatments with Dnmti post CCL_4 treatments gave an important insight into the role of DNA methylation in silencing the Tet driven transgene expression. However, certain questions were still unanswered. This concerned the question if methylation was established already on ES cell level or if it was established during development. Another question was if DNA methylation was the only factor contributing to silencing or if there were other epigenetic mechanisms contributing to silencing.

To investigate such issues the RosaGFP embryonic stem cell line was used which represents the earliest state directly after targeting. To evaluate the expression level, the RosaGFP mES cells were kept with (2 µg/ml) Doxycycline for a period of 72 hrs and subjected to flow cytometry analysis. Strikingly, the frequency of GFP expressing cells in the presence of Dox was only about 1.6%. To evaluate if this low number of GFP expressing cells could be rescued with the treatment with Dnmti, the Rosa-GFP mES cells were treated the Rosa-GFP mES cells with 0.5ug/ul Azacytidine and Decitabine for 72 hours. The cells were analyzed both independently and in combination. Then the cells werereanalysed the GFP expression through flow cytometry. Interestingly, the transgene expression in the mES cells increased to 6 and 8% upon treatment with Aza and Deci, respectively and nearly 10% following the treatment with the combination of both Dnmti (Fig 24). However, a full activation was not obtained. This suggests thatthe Dnmti can increase the expressionof the cassette in ES cells.

However, it cannot be excluded that the treatment was not fully efficient or that there might be other factors that might also contribute towards stabilizing the silencing of the Tet promoter apart from DNA methylation.

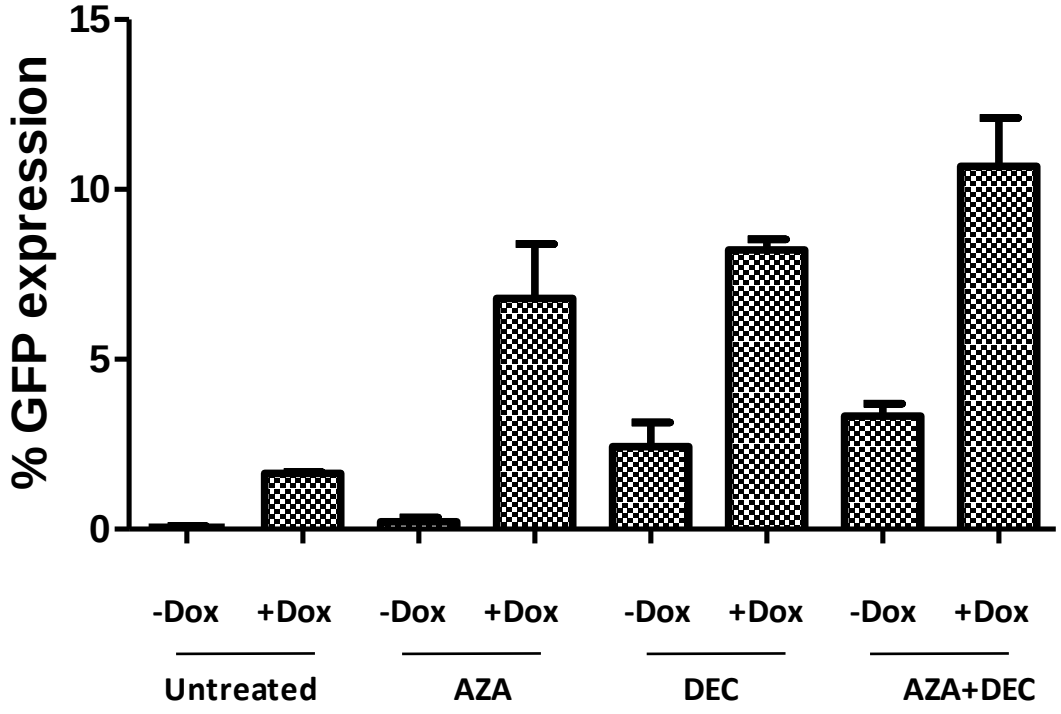


Figure 24: Treatment of the Rosa-GFP mES cells with Azacytidine (Aza) Decitabine (Deci) at a final concentration of 1μM and 0.5 μM respectively for 72 hrs. The graph shows partial rescue of GFP expression upon treatment with these inhibitors in a synergistic manner (bars represent three replicate per group).

Still, even without the Dnmti the percentage of GFP positive mES cells did not correspond to the lack of expression in the tissue (cf Figure 21). This suggests that expression is lost during development. To evaluate the developmental stage when the BiTet promoter was methylated, RosaGFPmES cells were differentiated *invitro*by formation of hanging drops. Interestingly, Tet driven transgene expression was significantly lost upon these early steps of differentiation (data not shown). The loss of transgene expression could be attributed to high levels of methylation of the Tet promoter right from early stages of differentiation (data not shown).

3.2.6 Cell to cell heterogeneity in the Tet methylation pattern of RosaGFP ES cells

The presence of both expressing and non-expressing cells in the same cell culture suggests cell to cell variability in the activation of the BiTet promoter in ES cells (Figure24). Such stochasticity can be due to epigenetically altered chromatin in individual cells which could contribute to the heterogeneous phenotype observed. In order to study the heterogeneous phenotype in more detail, theinduced RosaGFP mES cells were sorted upon treatment with the Dnmti Azacytidine and Decitabine (1µM and 0.5 µMfor 72 hrs).

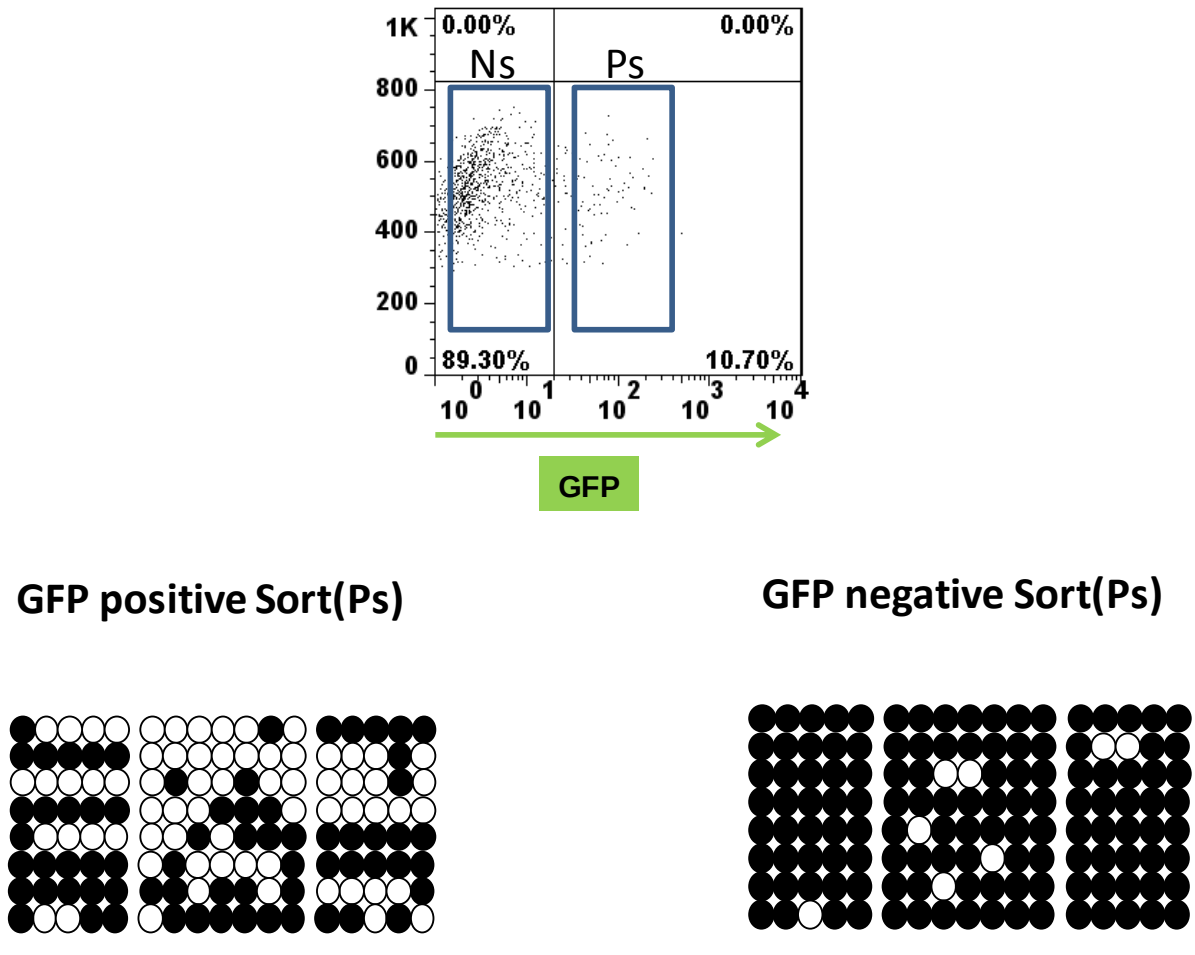


Figure 25: Flow cytometry plot of RosaGFP mES cells upon treatment with Azacytidine (Aza) and Decitabine (Deci) and the gating strategy used for sorting. DNA from the two sorted populations was isolated and was subjected to bisulfite conversion. Analysis was performed as for Figure 11

Interestingly, the methylation analysis of the GFP positive fraction revealed drastic reduction of the overall methylation levels to 50% while in the GFP negative fraction nearly 90% of mCpGs were observed. Together, this indicates a certain correlation between the transgene expression and BiTet promoter methylation

3.2.7 Evaluation of a CpG free bidirectional Tet promoter in the Rosa26 locus

The DNA methylation of bidirectional Tet promoter was shown to contribute to suppression of transgene expression. Though partial rescue was achieved upon treatment with the DNMT inhibitors the methylation analysis showed a significant level of methylation even in the GFP sorted mES cell population. This strongly suggests that there might be other mechanisms responsible for suppressing the Tet driven expression. To prove this, all the CpG motifs in the bidirectional Tet promoter were eliminated without changing the CpG motif of the coding regions. For this purpose, a modified, CpG free synthetic BiTet promoter was synthesized (Figure26) in which all the CpGs were replaced by TG. The promoter was cloned into the Rosa-GFP construct, thereby substituting the original BiTet promoter. Thereby the vector was obtained with both the GFP and luciferase under the control of the CpG free bidirectional Tet promoter.

First, the activity of the promoter was tested. For this purpose, the construct was transfected into HEK293T cells, together with reverse transactivator plasmid (rtTA). Two days after transfection, the cells were evaluated for GFP expression.

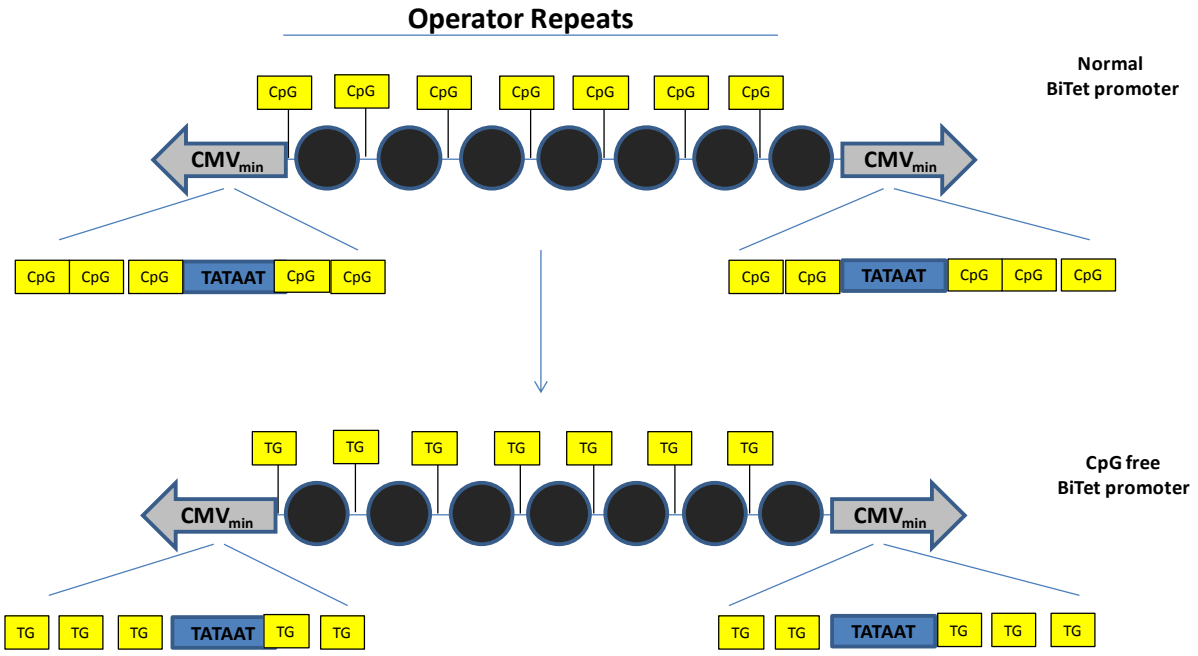
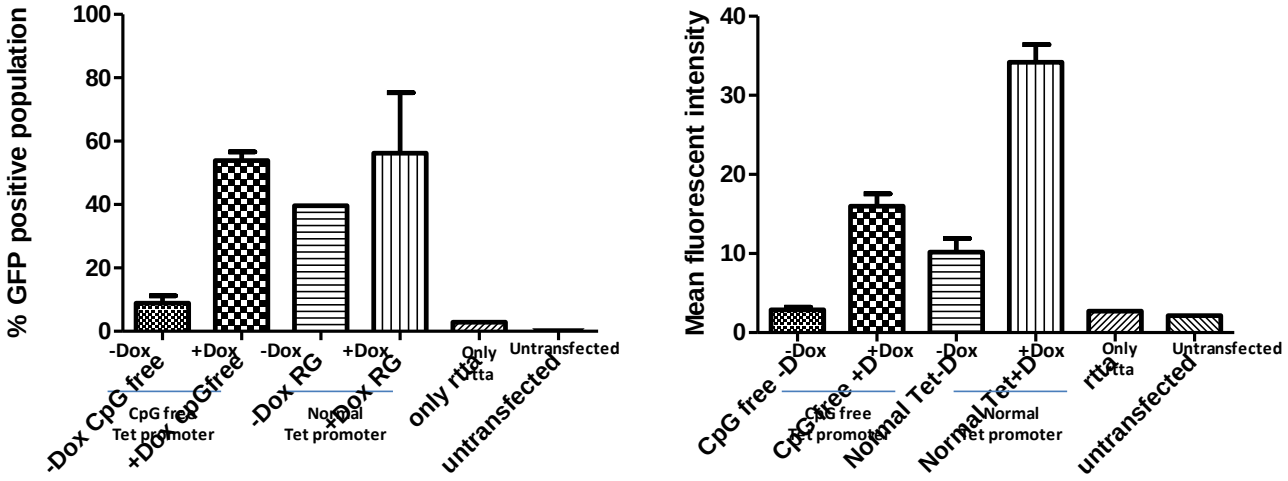


Figure 26: Depiction of the de novo synthesized CpG free Rosa GFP vector. CpG free Tet promoter was synthesized where the all the CpG motifs were replaced by TG.

A



B

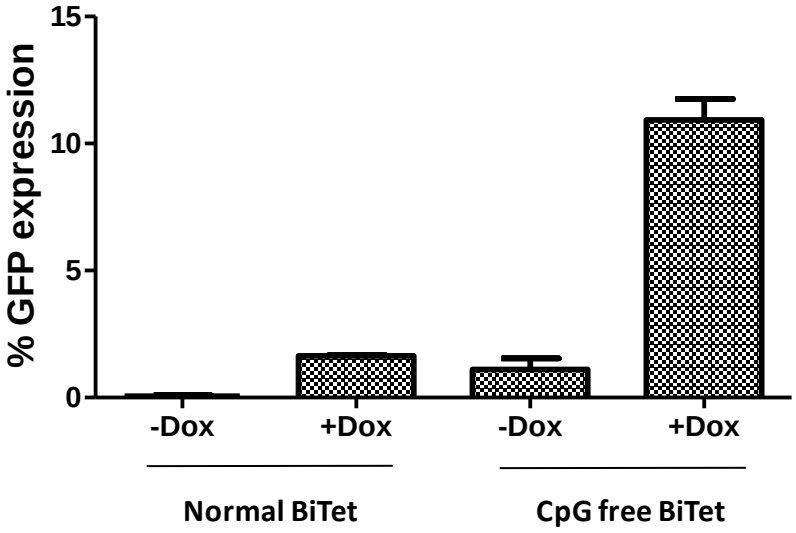


Figure 27: Expression analysis of GFP driven by normal and CpG free BiTet promoterA) GFP expression analysis upon transient co-transfection of CpG free Rosa GFP vector and normal Rosa-GFP vector in HEK293T cells along with rtTA. B) Upon integration into the Rosa 26 locus through RMCE, clonal population of ES cells were generated and analyzed for expression upon cultivation without and with Doxycycline.

Importantly, the CpG free promoter showed nicely regulated expression in the transient experimental settings in 293T. The overall expression level was comparable to the classical promoter. This indicates that elimination of the CpGs did not compromise the strength or inducibility.

In order to evaluate the performance of the CpG free promoter in the Rosa26 locus, the vector was targeted into the RMCE competent ES cells by Flp mediated RMCE according to the previously published protocol (99).It was expected that in these cells expression would not be compromised by DNA methylation mediated silencing. Targeted cells were analyzed for GFP expression in presence and absence of Doxycycline according to the previously described procedure. Strikingly, although all the potentially methylated CpG motifs have been eliminated only 7% GFP expressing mES cells were identified in induced state. Thus, a significant number of cells did not express the transgene.

Together this suggests that the Tet driven expression in the Rosa26 locus might not only be hindered by DNA methylation but that the BiTet promoter is also affected by other epigenetic mechanisms.

4 DISCUSSION

Transgenic models have given the researcher a new tool to carry out investigation concerning question in basic research. Whether it is model a disease or using a transgenic model for drug screening, the overall utility of transgenic models has been immense. Not only basic research but also the field of biotech have benefited from development of various recombinant models allowing expression of heterologous proteins from recombinant cell lines. However the major bottleneck in the transgenesis is the stability of expression of the transgene. With single cell analysis now available, the phenotypic heterogeneity in expression has now been reported frequently. This variable transgene expression is considered to be a result of the crosstalk of the incoming promoter cassette with cis-acting elements associated with the chromosomal site of transgene integration (position effect). However, the understanding of underlying mechanisms mediating these interactions of proximal as well as distal cis and trans acting regulatory elements has remained unclear. To have a mechanistic understanding of the nature of interaction that occurs upon the integration of thetransgene in defined chromosomal sites and how the transgene expression is modulated more than 100 independent cell clones in two different cell lines (CHOK1 and HEK293T) having single copy of hCMV promoter driven transgene were analyzed for variability in inter and intraclonal expression phenotype.This approach was based on a traditional screening approach where the various integration sites were screened for high expression and analysed for the transgene expression.

The study on CHOK1 and HEK293T cellscould shed light on the epigenetic modulation upon transgene integration occurring in two biotechnologically relevant cell lines. The findings of histone regulation of the CMV promoter involved in intraclonal variation provide novel insights into the nature of epigenetic crosstalk between the incoming cassette and the integration site. Unexpectedly, the heterogeneity associated with the hCMV driven constructs in the CHO and HEK293T clones was not associated with DNA methylation. In contrast, the hCMV promoter was associated with distinct histone modifications causing variable transgene expression and transgene silencing. Importantly, upon partially altering these histone marks with treatment with epigenetic modifiers like Sodium butyrate and Valproic acid, known *HDAC inhibitors,* transgene expression was partially restored. This

supports the idea that the histone modifications played a major part in activating or in suppressing transgene expression upon cell propagation in CHO and HEK293T cell.

Targeting of pre-selected chromosomal sites has been proposed to limit the position dependent expression of transgenes (137,138). The integration of a transgene in predefined sites allows for predictable expression since the effect of the neighbouring chromosomal elements are defined and thus the environment of the transgene doesnot change much. Thus, in the second approach of this thesis the known open chromatin locus Rosa26 in the murine system was targeted with Tetracycline inducible synthetic promoter (Tet) cassettes. Targeting was expected to give a predictable and homogenous transgene expression. Integration of synthetic constructs adds a new dimension to the already complex interaction that occurs when a transgene controlled by natural promoters is integrated in the genomic locus. The Tetracycline inducible promoter (Tet promoter system) is one such synthetic promoter element that has been designed to yield a controlled transcription of genes under it. This Tet promoter comprises of a minimal viral CMV promoter along with fused operator sequences from the bacterial operon system to which the transactivators binds and thereby activating transcription from the CMV promoter fragment fused to it. Thus, the combination of viral and bacterial elements in the Tet system can result in a unique crosstalk when integrated in the defined locus.

However, targeting of the Rosa26 locus with the synthetic cassettes resulted in a highly heterogeneous and low expression. The results clearly show that cassettes with Tetracycline (Tet) inducible promoter system are highly susceptible to epigenetic silencing by mechanisms like DNA methylation in particular.The transgene expression was evaluated upon single integration of Tet driven transgene in the Rosa26 locus that was previously suggested to support transgene expression(96,97,139). The result of this thesis shows that the general expectation that Tet driven transgene work efficiently in these known loci is not valid and robust for different experimental settings. The tendency of Tet promoters to get methylated and loose expression was shown to be the case in both the Rosa26 as well as Col1A1 locus. Interestingly, though the treatment with drugs like Azacytidine and Decitabine did lead to reduction in the DNA methylation levels and also lead to the partial recovery of the expression of the GFP and luciferase driven by the BiTet promoter, the recovery was not obtained in all cells. Rather, a highly heterogeneous cell to cell variation was observed.

4.1 Spreading of DNA methylation

The hypothesis of spreading of DNA methylation from a highly methylated foci to nearby regions was showed elegantly in a study by Jahner and Jaenish(140) were an integrated retroviral elements lead to de novo methylation of chromosomal sites nearby. Further, it was proposed by Tucker et al(136) that this spreading of DNA methylation is caused by the highly methylated regions that acts like a nucleation centre and triggers the spreading of DNA methylation to distal sites. This event is mechanistically driven by recruitment of repression complexes at the methylated sites. These complexes then interact with the nearby region and alter the chromatin making it more susceptible to DNA methylation machinery.

Thus, to evaluate if such kind of crosstalk also occur in the Rosa locus since the Tet promoter being highly methylated (>90% see Figure 22) could act as the nucleation centre triggering the methylation of nearby regions, the endogenous neighbouring regions having CpG island promoters namely the Rosa26 promoter and Thumpd3 promoter were evaluated for their methylation state. However, the Rosa26 and Thumpd3 promoter largely remained methylation free. This might be explained by the fact that there might be certain protective mechanism that protect these active promoter regions from the effect of nearby methylation(134).

4.2 Screening as a method to identify stable integration sites in cell lines and mice

Stable and predictable expression pattern are highly desirable properties in cases of generating genetically modified organisms with the use of foreign DNA, be it in the development of recombinant cell lines or transgenic animal models today. Till today, the biotech industry rely on random integration of the transgene before adopting large scale cumbersome screening measure to look for the high producer cells and then going for clonal selection. As analyzed in this study, this method not only required large number of clones to be screened but is also marred by instability of the clones due to position effects and epigenetic silencing.

In the first part of the study, the stability of transgene expression in HEK293T cells and CHOK1 cells was investigated. Large clone-to clone (interclonal) and cell-to-cell (intraclonal) variation was observed. Also, a considerable drop of transgene expression was observed during the early phase of establishment of clones. This was more pronounced in CHOK1 derived cell clones. The clones were established from cells that were initially identified according to their high GFP expression. However, a large variance in the GFP expression was observed when the clonal populations derived thereof were analysed at passage 2. Of note, this was observed both in the HEK293T cells as well as in CHOK1 cell derived clones, however to different extents. On the one hand, this variability concerned the clone-to-clone variation. This could be explained by the variability in chromosomal integration sites lending respective differences in the modulation of transgene expression (position effects). This modulation was observed during early phases of cultivation and gave rise to large differences of expression (Figure 9 and 10). Interestingly, also cell clones were identified with large intraclonal heterogeneity in transgene expression. To identify the underlying mechanism, subpopulations displaying high and low/no GFP expressing were isolated, respectively, from these heterogeneous clonal populations. It was observed that while in some clones these subpopulations stably maintained their phenotype over passaging, clone 31^T and 42^T showed a dynamic (metastable) phenotype with populations shifting between the expressing and non-expressing states (Figure 16). In these clones, the expressing and non-expressing subpopulations merged upon passaging. Thus, these clones represent a metastable state with continuous stochastic alterations in gene expression. This shift in the expression was highly interesting and might be explained the effect of transcriptional bursts modulating the expression of transgene. Transcriptional burst is a phenomenon where multiple transcripts are produced in short span of time due to combined firing of RNA polymerase II and the it is followed by a phase of non transcription(141,142). Since the lentiviral particles are known to integrate in a semi-random manner, with most of them integrating in the transcriptionally active region(141), heterogeneity in transcription of transgene could be conveyed by the site of integration (143).

Another reason for heterogeneity in transgene expression could be the stochastic inactivation that has been previously observed for genes located in the vicinity of a heterochromatic region. The spreading of these heterochromatin regions have been shown

to occur to a different extent and thus comprise a variable length in individual cells leading to the position effect variegation (PEV) (144-147). Similarly, transgenes inserted within or in the vicinity of heterochromatic regions often become stochastically inactivated (148). Thus, from the screening approach it was evident that most of the integration events occurred at sites in the genome that didnot support high expression of the transgene and are might also have been marred by possible modulation from the transcriptional burst. Therefore, most integration sites lead to large variability in the clonal properties.

4.3 Predictability of transgene expression upon targeted integration and its limitations

Recombinase mediated cassette exchange was employed on one hand to investigate if an expression pattern is re-established on a neutral cassette upon targeted integration of plasmid DNA in tagged CHOK1 and HEK293T cells.

Also, this method (RMCE) was used to answer if a defined, ubiquitously expressing integration site could be used to achieve conditional regulated expression from synthetic BiTet driven transgenes from pre tagged Rosa26 locus.

4.3.1 Targeted integration in preselected loci in HEK293T and CHOK1

To achieve this in CHOK1 and HEK293T, a bias-free approach was adopted and the targeted clones were sorted for the loss of fluorescent marker (GFP/RFP) that confirms the excision of the parental cassette. Importantly, this avoided any selection for expression of a resistance marker which might have imposed per se, alterations in the epigenetic pattern of the incoming cassette. Interestingly, by re-integration of the epigenetically neutral CMV promoter driving RFP as a reporter, the phenotype of the parental cells was re-established. This was also true for CMV based inducible Tet promoter driven expression. However, in contrast, upon targeting with the SV40 promoter driving the same reporter, the predictability in the expression was completely lost. Only one out of three cell clones could reproduce the expression status, while the others dropped in expression. Together, this indicates a differential influence of a given chromosomal integration site on incoming promoters. This might explain the previously reported finding that predictable expression upon targeted integration is restricted to the same design of expression cassettes (116,130).

4.3.2 Targeted integration in Rosa26 locus

Targeting of pre-selected chromosomal sites has been proposed to overcome position dependent expression of heterologous expression (108,117). Indeed, this method allows predicting expression of transgenes and thereby facilitates fast and reliable establishment of production cell lines (109,115,149). Another alternative to achieve predictable transgene expression is to look for loci which support stable expression and reuse them.Since a number of loci are already available/known supporting either stable constitutive transgene expression (150,151) and/or inducible expression of transgenes (139)(96,152-154) These genomic loci have been adapted for site-specific recombination technology that will allow the efficient reuse of these genomic loci. These strategies were employed/introduced to circumvent the need for doing large scale screens for every new transgene that needs to be expressed. With this thesis, the well known Rosa26 locus was targeted with the synthetic Tet promoter driven construct and subsequent transgene expression was evaluated. The Tet combination with these two loci was expected to show strictly controlled transgene expression. Contrary to the expectations, the Tet promoter was found to be highly prone to DNA methylation in the Rosa26 locus (Figure 22).The targeting of Rosa26 was expected to support transgene expression as it has been previously reported by various studies conducted in other laboratories(155). However, contrary to general expectations the predictability was compromised even in this case. The expression from the Tet promoter driven transgenes was heterogeneous with high mouse to mouse variability (PhD thesis, Natascha Kruse 2013 and other data not shown). This kind of variability was not restricted to different mouse but was even seen at the cellular levels (see Figure21)(139). These findings again emphasise that the settings adopted in screening in different studies with different readouts for transgene expression might lead to variable conclusions. However, the findings in the current thesis support the highly unstable and variable BiTet driven transgene expression in the Rosa26 locus

4.4 Epigenetic mechanisms involved intransgene silencing

The role of epigenetic mechanisms like DNA methylation and histone repression has been known to play a key role in the silencing of transgenes. Most often the transgene silencing has been shown to lead to a loss of transgene expression not only in cell lines based systems but also *in vivo* in mice. These epigenetic processes were known to act together to silence the transgene expression (156,157)especially in cases where transgene form tandem repeats(158). However the extent to which each of these mechanisms can act individually is still not clearly understood.

4.4.1 Evaluation of transgene modulation in HEK293T and CHOK1 cell lines

To identify the epigenetic mechanism underlying the unstable and heterogeneous transgene expression in CHO and 293T cell clones, a systematic evaluation was performedto study the impact of DNA methylation and histone modifications(Figure12 and 13). The study focused on clones with single copy integration events. Thereby, the heterogeneous and unstable expression in the selected clones could be attributed to a specific integration site and could exclude overlapping effects that would arise from multi-copy integrations. This excluded cassette induced silencing which was previously reported to accompany tandem or multi-copy integrations in defined loci. In this situation unstable transgene expression was observed which was independent on the copy-number (158-160).

Importantly, in the selected clones from HEK293T, the expression status of the clonal populations correlated with a specific histone modification pattern. Histone modifications have been considered to convey dynamic changes that can be triggered for e.g. by slight change in the environment (161,162). Decrease in histone H3 acetylation can lead to compaction of the DNA since these acetyl groups are known to neutralize the positive charges of histones and thus preventing strong interaction between histone and negatively charged phosphate group of DNA (163). Also, markings like H3K27me3 have been known to lend a stably silenced phenotype in DNA methylation independent manner. In agreement with this, a pronounced enrichment of Histone H3 lysine 4 acetylation (H3k4ac) marks in all

the positive sorted populations and the more frequent occurrence of the Histone H3 lysine 27 trimethylation (H3K27me3) in all the negative sorted populations were observed. The surprising factor was the absence of DNA methylation in the clonal populations that lost GFP expression stably both in 293T as well as in CHOK1 derived clonal populations (Figure11). DNA methylation state of CMV promoter was also evaluated at late passages (Figure11), however complete absence of methylation signified the stability of histone mediated repression. Also, the presence of other mechanism involved that might contribute to stabilize this repression apart from DNA methylation can also not be ruled out.

4.4.2 Epigenetic influences on Tet promoter in Rosa26 locus

The role of DNA methylation did not significantly contribute to the silencing of CMV driven transgene in CHO and HEK293T. However, the contribution of this epigenetic mark was significantly more prominent in the Tet driven transgene in the Rosa26 (Figure21 and 22). The targeting of transgenes in the open, ubiquitous loci like the Rosa26 was expected to give a predictable and homogenous transgene expression. However, the results (Figure22) clearly show that cassettes with Tetracycline (Tet) inducible promoter system are highly susceptible to epigenetic silencing by mechanisms like DNA methylation in particular. Interestingly, there have been already few indications that the Tet promoter is silenced in particular sites. It was previously published (164)that the BiTet promoter is susceptible to epigenetic modifications. This study was done in random integration settings and focused only on the brain tissue. The mechanism that might underlie crosstalk between the synthetic Tet driven constructs and defined sites in the chromosomes have largely remained out of focus. The current study looked into such kind of crosstalk and the underlying mechanism. The transgene expression was evaluated upon single integration of Tet driven transgene in the Rosa26 locus that is known to support transgene expression(96,97). The Tet promoter, although CMV derived also have parts from bacterial operon system and this might be the trigger that makes the Tet promoter susceptible to silencing by getting methylated.However the role of the integration site cannot be overlooked. Since the same Tet promoter worked very efficiently upon targeting in the 293T CL 17 PS (Figure 18) suggests that upon targeting the promoter in the favourable position in the genome, the crosstalk can support Tet driven transgene expression. However, the same promoter system

in different genomic location can be silenced upon unfavourable crosstalk. This was evident in case of Tet promoter upon targeting in the Rosa26.

4.5 The CMV promoter and its sensitivity towards silencing

The CMV promoter is one of the most commonly used promoters for transgene expression. However, inspite of the strong viral activity of the CMV promoter, it has been shown to be prone to epigenetic silencing. The role of DNA methylation in the silencing of the CMV driven transgene in CHOK1 and HEK293T was found to be non significant (Figure11). This is in contrast to the role of DNA methylation in CMV promoter silencing shown in the study by (119) et al where they have shown that CMV promoter is preferentially silenced in the mES cells. However, with the use of the DNMT knock out cells, the expression was not lost and the preferential silencing of CMV didnot take place emphasizing the key role of DNA methylation mediated silencing of the CMV promoter. However, this was study focused on the silencing of CMV promoter in mES cells in the episomal state and therefore the different experimental setting can explain the differences that were observed.

The role of histone markings on the CMV promoter and its effect on the transgene expression was studied in details by Mehta et al (120)where they found histone marks that could lead to promoter silencing. In the same study they found bivalent chromatin marks with CMV promoter being enriched with both repressive and active marks. These bivalent markings were held responsible for CMV promoter being in a state from where it can be activated. It is tempting to speculate that the metastable phenotype observed in $Cl31^{T}$ and $Cl42^{T}$might also have bivalent chromatin marks that might undergo dynamic changes causing switch in the phenotype (Figure12).

4.6 Potential factors responsible for transgene silencing

The transgene expression in eukaryotic cell has been hindered by silencing of the transgene. A number of characteristics have been associated with the transgenic elements that make them susceptible to silencing. These are 1) transgenes are integrated in array of tandem repeats 2) most of them have base sequences that is not common to eukaryotic environment 3) most often they are driven by strong viral promoters making them unusual

for the normal eukaryotic environment. These factors singly or in combination might trigger the cell to treat these transgenic elements similar to transposons or to retroviral elements and thereby might silence them.

Till date large number of factors has been postulated to trigger transgene silencing. The transgenic experiments in generally have involved transfection based approaches that have yielded large number of copies integration at a site in a tandem repeat manner. In a study by Manuelidis .L .*et al*(165) it was found that 11 megabase of transgene was inserted in the mouse chromosome in tandem repeat array and these repeats were enough to undergo heterochromatinization.

The sequence of transgene inserted might also act as a trigger for the transgene silencing in an effect that might be modulated by cis acting factors. Again this might not be exclusive from the fact that the multicopy silencing is more common then single copy. These cis acting factors might also be related to the CpG content. The CpG content has been shown to play a major role especially when present in the transcribed region and also it has been proposed that the threshold level of the CpGs that are considered normal exists in the genome exceeding which might also act as a trigger for suppression (166).

The presence of strong viral promoters such as the CMV promoter or viral LTRs are reported to be very prone to getting silenced (167).This again can partly be dependent on the sequence context. The prokaryotic genome has unmethylated CpG dinucleotideswhereas the eukaryotic genome has one-fifth of the CpG dinucleotides are methylated (Ref). These can be detected by the cell and can cause silencing as a protective mechanism adapted by the cell against the parasitic DNA(167).

4.7 The potential of epigenetic modifiers (Dnmtinhibitors and HDAC inhibitors) to alter chromatin state and recover transgene expression

It has been known that two kind of major epigenetic changes account for loss of transgene expression. One is DNA methylation and other is chromatin conformation alteration due to histone modification. Transgenes such as lentiviral SIN vectors carrying internal promoters such as EF1 alpha, CMV and CBA have been shown to undergo rapid silencing in the proviral state itself (Jin He et al,J V 2005).However, several reports show that the transgene

expression is recovered partially upon treatment with DNMT inhibitor azacytidine(168). Interestingly, also in case of the Tet promoter, the treatment with drugs like Azacytidine and Decitabine did lead to reduction in the DNA methylation levels and also lead to the partial recovery of the expression of the GFP and luciferase marker genes driven by Tet. In most cases, the DNA methylation leads to recruitment of methyl binding proteins that further recruit other modifiers like Histone deacetylases (HDACs). Thus, the rescue mostly is not hundred percent upon treatment with DNMTi drugs like Azacytidine. This suggests that there might be other mechanisms that might be involved in orchestrating the silencing of Tet driven transgenes.Therefore, invariably in most cases the methylation of cytosine is also accompanied by modification of histone marks and this is consistent with the observation that the treatment with Azacytidine lead to partial recovery and this is lost overtime and cell stop responding to Azacytidine. However periodic treatments with combination of these drugs have been found to be able to sustain transgene expression (Kong Q, 2011) for a period of more than 60 days.

The HDACi inhibitors can remove HDACs and prevent hypoacetylation that has been known to cause compaction of DNA. Inspite of all this literature it was highly interesting that the clones characterized in depth did not show any significant contribution of DNA methylation (Figure 11). However, the effects from HDAC treatment did indicate the loss of acetylation in the GFP non expressing cells thus suggesting that histone mediated repression not only can occur in the absence of DNA methylation but can also be have stable repression. The non-expressing populations of clones 31^T and 42^T increased transgene expression in response to the HDACi; Valproic acid (VPA) and Sodium butyrate (NaB) (Figure13). In contrast, non-expressing populations of clone 12^T, 17^T 54^T did not respond to this treatment (data not shown). This indicates that these clones are stably silenced (48).

These results might be explained by bivalent histone markings. In this case two kinds of histone marks are present, one being repressive marking and other being active marking. In such case the bivalent region is said to be poised and can be reactivated by treatments with HDACi inhibitors once the repressive modifications are removed. However in certain regions of the genome, these histone repressive markings can be catalysed by different kinds of enzymes. Hst1 is a HDAC from class 1 HDAC family that contributes to deacetylation from the euchromatin region and is therefore susceptible to HDACi like NaB, TSA and VPA (J R

Dave, 2003). However, Sir2 belonging to class III of histone deacetylases, and are known to cause deacetylation of heterochromatin regions and these are not susceptible to HDACi like VPA, NaB and TSA, therefore once a transgene undergoes heterochromatization the reactivation by HDACi treatments can no longer take place.

4.8 Tetracycline inducible promoter in Rosa26 locus: The Rosa-Tet combination

The inducible expression from Tet promoter system has been a revelation on its own making possible otherwise difficult experimental settings such as conditional expression of toxic compounds and developmentally controlled genes (169-171). The targeted integration of Tet promoter in Cl17 of HEK293T confirmed that the Tet system integrated in the single copy can give strong regulation and high expression. However, the importance of the Rosa26 as an integration site for Tet promoter based transgene expression lies in generation of transgenic mice since it is known to express during developmental stages and give a uniform expression pattern. The work from Masui et al ( ) developed the Rosa Tet system where they successfully targeted the mES cells with Tet promoter in the Rosa26 locus and could achieve successfully the regulated expression. This study combined the ubiquitously expressing Rosa26 locus with the Tet promoter based system. Although they were successful in achieving Tet driven expression, they observed that the strict regulation was lost after five passages and also the transgene expression was variable. The study discussed the possibility of transgene silencing and suggested that the effect could be the result of specific sequences of coding genes integrated under the control of Tet promoter. In the present study, affect from a different coding sequence was also analyzed. This was done on Rosa T-antigen mouse model in which the BiTet promoter was used to express SV40 T antigen (data not shown). There too, both in vitro in mES cells and in vivo, it was found that the transgene expression from BiTet was hindered dues to DNA methylation.

Other studies have been conducted in mice showing the utilization of the Tet promoter based system upon targeting in the Rosa 26 locus(138,139,155). These studies were based on expression constructs that were slightly different designs with respect to orientation, intervening sequences, and targeted gene. While the Dox dependent expression was described in the Rosa locus, work from Strathdee et al (155) suggested that transcriptional

interference can cause problems and therefore the Tet promoter based expression depends on the orientation of the promoter with respect to endogenous Rosa promoter. This conclusion seems to be contradictory to the results shown in this thesis. Indeed, it was found that both the bidirectional Tet promoter (BiTet) and the unidirectional Tet promoter (in both of the possible orientations) could not give a predictable and uniform expression (data not shown and PhD thesis, Natascha Kruse, 2013). This suggests that orientation dependency is not the major cause of heterogeneous and unstable Tet driven expression. Rather, the differences to the other studies might be explained by different read outs. The work from Strathdee et al mainly focused on mES cells and overlooked the epigenetic mechanism resulting in the differences in the expression phenotype as is reported in the current thesis work.

In light of the results presented in the result section (chapters 3.2.1-3.2.3), and some of the work from other groups discussed above, the general proposition that Tet driven transgene expresses efficiently in Rosa26 loci cannot be considered to be valid and robust for different experimental settings. The tendency of Tet promoters to get methylated and loose expression was shown to be quite significant even in the ubiquitous Rosa26 locus.

4.9 Procollagen (Col1A1) locus as an alternative to Rosa26?

The problems encountered during expression from the Tet driven transgenes in Rosa26 locus lead to search for other loci. One such locus was found to be the Col1A1 locus. The Procollagen locus (Col1A1) was first targeted in sheep(172)It was shown that this locus was able to support transgene expression. This eventually lead to the identification and targeting of this locus even in mouse(96). In particular, the site was exploited for targeting the dox inducible Tet promoter into the ColA1 locus in mouse ES cells and was able to generate mouse which successfully expressed the transgene. Strikingly, inspite of the reported strict regulation with uniform expression(96), in the current study it was observed that the Tet driven transgene expression was found to be silenced in the mES cells upon differentiation (data not shown). The methylation analysis showed high levels of methylation during the early differentiation stages upon in vitro differentiation. This is partly in line with another recent study from Wan et al, 2013 (173) who showed that the expression from Tet promoter was sensitive to epigenetic influences in the Col1A1 locus. They showed that this

was related to the exposure to the Doxycycline during the early developmental stages. The conclusion from the study was that the Tet promoter can be silenced if the transcription from Tet is induced early on in developmental stage. But the silencing can be prevented and strict regulation can be achieved if the initial few days during embryonic growth does not involve the treatment of Doxycycline. Still, these conclusions are not clear since they imply that the chromatin conformation later or in cells might be dependent on initial transcription from Tet promoter. This is also contradictory to other results from other studies () where initial Dox treatments were held essential to switch on the Tet promoter.

4.10 CpG free Tet promoter- A potential solution to overcome silencing

With high levels of DNA methylation in the bidirectional Tet promoter, one possible way to circumvent the problem of DNA methylation could be to eliminate the CpG motifs present in the Tet promoter. Thus, in this study, the cytosines of the CpG motifs were replaced by thymine and thereby resulting in the CpG free BiTet promoter element. The promoter activity of this modified CpG free version of BiTet was found to work very efficiently and confers strict regulation. Upon targeting this CpG free Tet promoter in the Rosa26 locus, the expression was improved, however, the overall expression level was induced only by seven folds. This could be due the fact that change in the thymine from cytosine might have also affected the topology of this DNA fragment upon integration resulting in condensed state. Another explanation could be a potential role and significance of histone mediated repression and non CpG DNA methylation(174).

Also, with these results, it might be speculated that these silencing mechanism might act in layers with DNA methylation being one of the mechanism that might play a role to initiate silencing and might explain the partial rescue upon treatment with the DNMT inhibitors. However, the real cause of silencing of the Tet promoter in the Rosa 26 locus might not be attributed to a single mechanism but might be accounted by various different factors.

4.11 Challenge to the concept of "Safe Harbors"

The promise of genome engineering with stable integration of functional transgenes for the therapeutic purposes has also lead to additional safety concerns upon genetic disruption caused due to transgene integration. With the predilection of most of retroviral and lentiviral based vectors to integrate near the transcribed genes; the major threat was if such an integration event leads to activation of oncogenes. However, with nearly eight percent of human genome harbouring retroviral elements that stably integrated in the mammalian lineage during evolution(175), the idea that there are sites that can support these elements lead to the search for the so-called 'safe harbor' in the genome, especially with respect to human genome. Theoretically, the "genomic safe harbors (GSH)" would be those sites in the genome in which transgene can be integrated without disturbing the activity of endogenous genes and thus without having any deleterious consequences (oncogenes activation or disruption of tumor suppressor), and yet supporting a high and stable transgene expression. Several attempts have been made to define such safe harbor sites by criteria such as a minimal distance to standard genes and to tumor promoting genes (176,177). However, such definitions do not consider the topology of DNA, the possibility of inter-chromosomal crosstalk (178) and also the eventual perturbation induced by the integrated expression cassette.

While none of the know sites in human genome fulfils all the criteria of GSH, some sites like AAVS1, CCR5 and Rosa26 in humans have been proposed to be close to GSH. However, the studies with full validation for these sites in terms of transgene expression, safety to host upon integration are lacking(179). Although the integration of transgene in the Rosa26 locus did not result in the pathophysiology in with mES cells retaining full pluripotency and giving rise to a normal mouse, as shown in the studies. Still the study shows that the Rosa26 locus in mouse cannot be considered as a safe harbour given the fact the integrated cassettes (at least the Tet promoter cassettes) represent a target of pronounced epigenetic modification. In addition it was also shown that the expression in this locus varies depending on the stage of development (data not shown).

The work discussed in above was aimed to characterize the epigenetic modulation underlyingcrosstalk between transgene and integration sites. These understandings will not only helps to surpass limitations of transgene silencing but also will greatly increase the freedom of designing custom-tailored de novo functional transgenes that yield predictable and desirable expression.

5 Conclusion

While DNA methylation was the major player in the silencing of the Tetracycline based promoter systems in murine models, the heterogeneity associated with the hCMV driven constructs in the CHO and HEK293T clones was not associated with DNA methylation. In the cell lines, the hCMV promoter was associated with distinct histones modifications causing variable transgene expression and transgene silencing. Here, the histone modifications played a major part in activating or in suppressing transgene expression upon cell propagation in CHO and HEK293T cell.

The first part of the study looked into the epigenetic modulation among different integration sites in two different cell lines. The major focus was on elucidating the nature of interaction between integration sites and the integrated transgenes. The presence of different kinds of clonal populations was attributed to position effects due to different nature of integration sites. However, the intraclonal variation observed in metastable clones was found to be correlated with differential histone marks. These might be caused due to modulations from transcriptional burst resulting in high cell to cell variability in transgene expression.

The Tetracycline based inducible promoter system was found to be highly susceptible to silencing by epigenetic modifications in the second part of the thesis. The surprising aspect was the fact that inspite of being targeted in the ubiquitously expressing Rosa26 locus, still the BiTet were silenced. The findings in the present study not only undermines the utility of Rosa locus for mouse transgenesis incase of development of inducible mouse models but also brings up the mechanism that are involved in the unfavourable crosstalk that occurs in this locus upon targeting of synthetic BiTet promoter driven constructs.

It is proposed that while the heterogeneous expression is associated with different chromatin states, as conferred by various epigenetic role players (DNA methylation and Histone modifications), the stability and extent of the variation in transgene expression may largely depend on the nature of crosstalk between the chromosomal integration site and the synthetic construct which might be stochastic in nature, with some degree of cell to cell variability and involves epigenetic role-plays.

6 Outlook

The work described in this thesis contributes to the understanding of the underlying crosstalk between transgenes and define integration sites. The CHOKI and HEK293T cell clones showed high levels variability in transgene expression and also only few specific sites being able to support expression from heterologous promoter in the sites screened for high expression using a different promoter element. Thus the sites that initially might have open chromatin conformation along with transcriptional activating epigenetic signatures can undergo changes upon integration of heterologous promoter element making the site unfit for high expression. This understanding can be useful for designing a strategy to screen for integration sites and these sites might be promoter specific.

The implication of silencing of BiTet promoter in the Rosa26 locus has shown that inspite of this locus being held as the most reliable locus for mouse transgenesis, this locus is not fit for expression of transgene from synthetic BiTet promoter which is heavily marred by silencing and epigenetic modulations.

Thus, more options have to be considered for transgenesis which may include redesigning of transgenic cassettes with chromosomal elements like LCRS, insulators that can prevent the modulation of transgene expression from the neighbouring regions in the genome. The understanding and the characterization of underlying mechanism of the crosstalk in the present work will helps to circumvent limitations in transgene expression and greatly expands the use of transgenic tools in investigating important questions both in basic research and biotech.

7 Summary

In an attempt to understand the underlying mechanism of the crosstalk that occurs between the transgene and chromosomal locus where the transgene gets integrated, the investigation of the epigenetic role play was the focus of the study. The epigenetic mechanism(s) were evaluated in CHOK1 and HEK293T cells where inter and intraclonal clonal heterogeneity of expression upon single copy transgene integration into random chromosomal sites was observed. These variations were reflected by differential histone modification pattern. We not only identified clones in which the modifications set in early after genetic modification and were stably inherited during cell division but also clones in which these modifications were dynamic and occurred stochastically upon extended cultivation. Importantly, we could restore a stable expression phenotype by targeting the chromosomal loci with a related expression cassette that re-established the favourable chromatin status. The investigation was also done to evaluate the suitability of Rosa26 locus for synthetic Tet promoter driven transgene expression. Inspite of the Rosa 26 locus being ubiquitously expressing; representing an open chromatin state, the BiTet driven constructs didnot shown stable expression in the Rosa26 locus and were heavily modulated by epigenetic markings in particular; DNA methylation.

8 REFERENCES

1. Penner, M.R., Roth, T.L., Barnes, C.A. and Sweatt, J.D. (2010) An epigenetic hypothesis of aging-related cognitive dysfunction. *Front Aging Neurosci*, 2, 9.
2. Osterlehner, A., Simmeth, S. and Gopfert, U. (2011) Promoter methylation and transgene copy numbers predict unstable protein production in recombinant Chinese hamster ovary cell lines. *Biotechnol Bioeng*, 108, 2670-2681.
3. Pauler, F.M., Sloane, M.A., Huang, R., Regha, K., Koerner, M.V., Tamir, I., Sommer, A., Aszodi, A., Jenuwein, T. and Barlow, D.P. (2009) H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Res*, 19, 221-233.
4. Schubeler, D. Molecular biology. (2012) Epigenetic islands in a genetic ocean. *Science*, 338, 756-757.
5. Schubeler, D. (2009) Epigenomics: Methylation matters. *Nature*, 462, 296-297.
6. Tiwari, V.K., McGarvey, K.M., Licchesi, J.D., Ohm, J.E., Herman, J.G., Schubeler, D. and Baylin, S.B. (2008) PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol*, 6, 2911-2927.
7. Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*, 9, 465-476.
8. Baubec, T. and Schubeler, D. (2014) Genomic patterns and context specific interpretation of DNA methylation. *Curr Opin Genet Dev*, 25C, 85-92.
9. Chen, R.A., Stempor, P., Down, T.A., Zeiser, E., Feuer, S.K. and Ahringer, J. (2014) Extreme HOT regions are CpG dense promoters in C. elegans and humans. *Genome Res*.
10. Gardiner-Garden, M. and Frommer, M. (1987) CpG islands in vertebrate genomes. *J Mol Biol*, 196, 261-282.
11. McGaughey, D.M., Ozel Abaan, H., Miller, R.M., Kropp, P.A. and Brody, L.C. (2014) Genomics of CpG Methylation in Developing and Developed Zebrafish. *G3 (Bethesda)*.
12. Guibert, S., Forne, T. and Weber, M. (2009) Dynamic regulation of DNA methylation during mammalian development. *Epigenomics*, 1, 81-98.
13. Uh, K.J., Park, C.H., Choi, K.H., Park, J.K., Jeong, Y.W., Roh, S., Hyun, S.H., Shin, T., Lee, C.K. and Hwang, W.S. (2014) Analysis of imprinted IGF2/H19 gene methylation and expression in normal fertilized and parthenogenetic embryonic stem cells of pigs. *Anim Reprod Sci*, 147, 47-55.
14. Patten, M.M., Ross, L., Curley, J.P., Queller, D.C., Bonduriansky, R. and Wolf, J.B. (2014) The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity (Edinb)*.
15. Joo, J.E., Novakovic, B., Cruickshank, M., Doyle, L.W., Craig, J.M. and Saffery, R. (2014) Human active X-specific DNA methylation events showing stability across time and tissues. *Eur J Hum Genet*.
16. Huh, I., Zeng, J., Park, T. and Yi, S.V. (2013) DNA methylation and transcriptional noise. *Epigenetics Chromatin*, 6, 9.
17. Subramaniam, D., Thombre, R., Dhar, A. and Anant, S. DNA Methyltransferases: (2014) A Novel Target for Prevention and Therapy. *Front Oncol*, 4, 80.
18. Yoder, J.A., Soman, N.S., Verdine, G.L. and Bestor, T.H. (1997) DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *J Mol Biol*, 270, 385-395.
19. Mohan, K.N. and Chaillet, J.R. (2013) Cell and molecular biology of DNA methyltransferase 1. *Int Rev Cell Mol Biol*, 306, 1-42.
20. Shanmugam, R., Aklujkar, M., Schafer, M., Reinhardt, R., Nickel, O., Reuter, G., Lovley, D.R., Ehrenhofer-Murray, A., Nellen, W., Ankri, S. *et al.*(2014) The Dnmt2 RNA

methyltransferase homolog of Geobacter sulfurreducens specifically methylates tRNA-Glu. *Nucleic Acids Res*.

21. Durdevic, Z., Mobin, M.B., Hanna, K., Lyko, F. and Schaefer, M. (2013) The RNA methyltransferase Dnmt2 is required for efficient Dicer-2-dependent siRNA pathway activity in Drosophila. *Cell Rep*, 4, 931-937.

22. Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99, 247-257.

23. Ge, Y.Z., Pu, M.T., Gowher, H., Wu, H.P., Ding, J.P., Jeltsch, A. and Xu, G.L. (2004) Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J Biol Chem*, 279, 25447-25454.

24. Gowher, H., Liebert, K., Hermann, A., Xu, G. and Jeltsch, A. (2005) Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *J Biol Chem*, 280, 13341-13348.

25. Koukoura, O., Spandidos, D.A., Daponte, A. and Sifakis, S. (2014) DNA methylation profiles in ovarian cancer: Implication in diagnosis and therapy (Review). *Mol Med Rep*.

26. Lubbert, M. (2000) DNA methylation inhibitors in the treatment of leukemias, myelodysplastic syndromes and hemoglobinopathies: clinical results and possible mechanisms of action. *Curr Top Microbiol Immunol*, 249, 135-164.

27. Santini, V., Kantarjian, H.M. and Issa, J.P. (2001) Changes in DNA methylation in neoplasia: pathophysiology and therapeutic implications. *Ann Intern Med*, 134, 573-586.

28. Day, J.J. and Sweatt, J.D. (2010) DNA methylation and memory formation. *Nat Neurosci*, 13, 1319-1323.

29. Zemach, A. and Zilberman, D. (2010) Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr Biol*, 20, R780-785.

30. Zemach, A., McDaniel, I.E., Silva, P. and Zilberman, D. (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, 328, 916-919.

31. Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E. *et al.*(2010) Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A*, 107, 8689-8694.

32. Messerschmidt, D.M., Knowles, B.B. and Solter, D. (2014) DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev*, 28, 812-828.

33. Li, E., Bestor, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69, 915-926.

34. Dodge, J.E., Okano, M., Dick, F., Tsujimoto, N., Chen, T., Wang, S., Ueda, Y., Dyson, N. and Li, E. (2005) Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *J Biol Chem*, 280, 17986-17991.

35. Medina-Jaime, A.D., Reyes-Vargas, F., Martinez-Gaytan, V., Zambrano-Galvan, G., Portillo-Delcampo, E., Burciaga-Nava, J.A., Reyes-Romero, M. and Sifuentes-Alvarez, A. (2014) ESR1 and PGR Gene Promoter Methylation and Correlations with Estrogen and Progesterone Receptors in Ductal and Lobular Breast Cancer. *Asian Pac J Cancer Prev*, 15, 3041-3044.

36. Brooks, W. (2014) A commentary on types of DNA methylation status of the interspersed repetitive sequences for LINE-1, Alu, HERV-E and HERV-K in the neutrophils from systemic lupus erythematosus patients and healthy controls. *J Hum Genet*, 59, 174-175.

37. Guo, L., Byun, H.M., Zhong, J., Motta, V., Barupal, J., Zheng, Y., Dou, C., Zhang, F., McCracken, J.P., Diaz, A. *et al.*(2014) Effects of short-term exposure to inhalable particulate matter on DNA methylation of tandem repeats. *Environ Mol Mutagen*, 55, 322-335.

38. **Bird, A. and Tweedie, S. (1995) Transcriptional noise and the evolution of gene number. *Philos Trans R Soc Lond B Biol Sci*, 349, 249-253.**
39. **Bird, A.P. (1995) Gene number, noise reduction and biological complexity. *Trends Genet*, 11, 94-100.**
40. **Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y. and Tyagi, S. (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*, 4, e309.**
41. **Golding, I., Paulsson, J., Zawilski, S.M. and Cox, E.C. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell*, 123, 1025-1036.**
42. **Biterge, B. and Schneider, R. (2014) Histone variants: key players of chromatin. *Cell Tissue Res*.**
43. **Talbert, P.B. and Henikoff, S. (2010) Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol*, 11, 264-275.**
44. **Balakrishnan, L. and Milavetz, B. (2010) Decoding the histone H4 lysine 20 methylation mark. *Crit Rev Biochem Mol Biol*, 45, 440-452.**
45. **Murr, R. (2010) Interplay between different epigenetic modifications and mechanisms. *Adv Genet*, 70, 101-141.**
46. **Garcia-Dominguez, M. and Reyes, J.C. (2009) SUMO association with repressor complexes, emerging routes for transcriptional control. *Biochim Biophys Acta*, 1789, 451-459.**
47. **Li, J., Lin, Q., Yoon, H.G., Huang, Z.Q., Strahl, B.D., Allis, C.D. and Wong, J. (2002) Involvement of histone methylation and phosphorylation in regulation of transcription by thyroid hormone receptor. *Mol Cell Biol*, 22, 5688-5697.**
48. **Jorgensen, S., Schotta, G. and Sorensen, C.S. (2013) Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic Acids Res*, 41, 2797-2806.**
49. **Shechter, D., Nicklay, J.J., Chitta, R.K., Shabanowitz, J., Hunt, D.F. and Allis, C.D. (2009) Analysis of histones in Xenopus laevis. I. A distinct index of enriched variants and modifications exists in each cell type and is remodeled during developmental transitions. *J Biol Chem*, 284, 1064-1074.**
50. **Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D. *et al.* (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406, 593-599.**
51. **Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature*, 403, 41-45.**
52. **Gardner, K.E., Allis, C.D. and Strahl, B.D. (2011) Operating on chromatin, a colorful language where context matters. *J Mol Biol*, 409, 36-46.**
53. **Briggs, S.D., Xiao, T., Sun, Z.W., Caldwell, J.A., Shabanowitz, J., Hunt, D.F., Allis, C.D. and Strahl, B.D. (2002) Gene silencing: trans-histone regulatory pathway in chromatin. *Nature*, 418, 498.**
54. **Sun, Z.W. and Allis, C.D. (2002) Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature*, 418, 104-108.**
55. **McCabe, M.T., Brandes, J.C. and Vertino, P.M. (2009) Cancer DNA methylation: molecular mechanisms and clinical implications. *Clin Cancer Res*, 15, 3927-3937.**
56. **Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T. and Peters, A.H. (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol*, 13, 1192-1200.**
57. **Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S. and Jacobsen, S.E. (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science*, 317, 1760-1764.**
58. **Delcuve, G.P., Khan, D.H. and Davie, J.R. (2012) Targeting class I histone deacetylases in cancer therapy. *Expert Opin Ther Targets*, 17, 29-41.**
59. **Delcuve, G.P., Khan, D.H. and Davie, J.R. (2012) Roles of histone deacetylases in epigenetic regulation: emerging paradigms from studies with inhibitors. *Clin Epigenetics*, 4, 5.**

60. **de Ruijter, A.J., van Gennip, A.H., Caron, H.N., Kemp, S. and van Kuilenburg, A.B. (2003) Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem J*, 370, 737-749.**

61. **Sengupta, N. and Seto, E. (2004) Regulation of histone deacetylase activities. *J Cell Biochem*, 93, 57-67.**

62. **Madsen, A.S., Kristensen, H.M., Lanz, G. and Olsen, C.A. (2014) The effect of various zinc binding groups on inhibition of histone deacetylases 1-11. *ChemMedChem*, 9, 614-626.**

63. **Gottlicher, M. (2004) Valproic acid: an old drug newly discovered as inhibitor of histone deacetylases. *Ann Hematol*, 83 Suppl 1, S91-92.**

64. **Kumar, P., Periyasamy, R., Das, S., Neerukonda, S., Mani, I. and Pandey, K.N. (2014) All-trans retinoic Acid and sodium butyrate enhance natriuretic Peptide receptor a gene transcription: role of histone modification. *Mol Pharmacol*, 85, 946-957.**

65. **Bantscheff, M., Hopf, C., Savitski, M.M., Dittmann, A., Grandi, P., Michon, A.M., Schlegl, J., Abraham, Y., Becher, I., Bergamini, G. *et al.*(2011) Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat Biotechnol*, 29, 255-265.**

66. **Smith, K.T., Martin-Brown, S.A., Florens, L., Washburn, M.P. and Workman, J.L. (2010) Deacetylase inhibitors dissociate the histone-targeting ING2 subunit from the Sin3 complex. *Chem Biol*, 17, 65-74.**

67. **Batenchuk, C., St-Pierre, S., Tepliakova, L., Adiga, S., Szuto, A., Kabbani, N., Bell, J.C., Baetz, K. and Kaern, M. (2011) Chromosomal position effects are linked to sir2-mediated variation in transcriptional burst size. *Biophys J*, 100, L56-58.**

68. **Chen, M., Licon, K., Otsuka, R., Pillus, L. and Ideker, T. (2013) Decoupling epigenetic and genetic effects through systematic analysis of gene position. *Cell Rep*, 3, 128-137.**

69. **Daboussi, F., Zaslavskiy, M., Poirot, L., Loperfido, M., Gouble, A., Guyot, V., Leduc, S., Galetto, R., Grizot, S., Oficjalska, D. *et al.*(2012) Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. *Nucleic Acids Res*, 40, 6367-6379.**

70. **Moreno, R., Martinez, I., Petriz, J., Gonzalez, J.R., Gratacos, E. and Aran, J.M. (2009) Boundary sequences stabilize transgene expression from subtle position effects in retroviral vectors. *Blood Cells Mol Dis*, 43, 214-220.**

71. **Yin, Z., Kong, Q.R., Zhao, Z.P., Wu, M.L., Mu, Y.S., Hu, K. and Liu, Z.H. (2012) Position effect variegation and epigenetic modification of a transgene in a pig model. *Genet Mol Res*, 11, 355-369.**

72. **Duan, B., Cheng, L., Gao, Y., Yin, F.X., Su, G.H., Shen, Q.Y., Liu, K., Hu, X., Liu, X. and Li, G.P. (2012) Silencing of fat-1 transgene expression in sheep may result from hypermethylation of its driven cytomegalovirus (CMV) promoter. *Theriogenology*, 78, 793-802.**

73. **Kong, Q., Wu, M., Wang, Z., Zhang, X., Li, L., Liu, X., Mu, Y. and Liu, Z. (2011) Effect of trichostatin A and 5-Aza-2'-deoxycytidine on transgene reactivation and epigenetic modification in transgenic pig fibroblast cells. *Mol Cell Biochem*, 355, 157-165.**

74. **Paredes, V., Park, J.S., Jeong, Y., Yoon, J. and Baek, K. (2013) Unstable expression of recombinant antibody during long-term culture of CHO cells is accompanied by histone H3 hypoacetylation. *Biotechnol Lett*, 35, 987-993.**

75. **Choi, K.H., Park, J.K., Kim, H.S., Uh, K.J., Son, D.C. and Lee, C.K. (2013) Epigenetic changes of lentiviral transgenes in porcine stem cells derived from embryonic origin. *PLoS One*, 8, e72184.**

76. **Simari, R.D., Yang, Z.Y., Ling, X., Stephan, D., Perkins, N.D., Nabel, G.J. and Nabel, E.G. (1998) Requirements for enhanced transgene expression by untranslated sequences from the human cytomegalovirus immediate-early gene. *Mol Med*, 4, 700-706.**

77. **Rotondaro, L., Mele, A. and Rovera, G. (1996) Efficiency of different viral promoters in directing gene expression in mammalian cells: effect of 3'-untranslated sequences. *Gene*, 168, 195-198.**

78. **Wen, S., Zhang, H., Li, Y., Wang, N., Zhang, W., Yang, K., Wu, N., Chen, X., Deng, F., Liao, Z. *et al.*(2014) Characterization of Constitutive Promoters for piggyBac Transposon-Mediated Stable Transgene Expression in Mesenchymal Stem Cells (MSCs). *PLoS One*, 9, e94397.**
79. **Cheng, L., Ziegelhoffer, P.R. and Yang, N.S. (1993) In vivo promoter activity and transgene expression in mammalian somatic tissues evaluated by using particle bombardment. *Proc Natl Acad Sci U S A*, 90, 4455-4459.**
80. **Gossen, M. and Bujard, H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A*, 89, 5547-5551.**
81. **Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W. and Bujard, H. (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science*, 268, 1766-1769.**
82. **Baron, U., Freundlieb, S., Gossen, M. and Bujard, H. (1995) Co-regulation of two gene activities by tetracycline via a bidirectional promoter. *Nucleic Acids Res*, 23, 3605-3606.**
83. **Loew, R., Heinz, N., Hampf, M., Bujard, H. and Gossen, M. (2010) Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol*, 10, 81.**
84. **Hwang, D.Y., Cho, J.S., Chae, K.R., Kang, T.S., Hwang, J.H., Lim, C.H., Lee, S.H., Lim, H.J., Min, S.H., Sheen, Y.Y. *et al.* (2003) Differential expression of the tetracycline-controlled transactivator-driven human CYP1B1 gene in double-transgenic mice is due to androgens: application for detecting androgens and antiandrogens. *Arch Biochem Biophys*, 415, 137-145.**
85. **Hwang, D.Y., Cho, J.S., Oh, J.H., Shim, S.B., Jee, S.W., Lee, S.H., Seo, S.J., Kang, H.G., Sheen, Y.Y. and Kim, Y.K. (2005) An in vivo bioassay for detecting antiandrogens using humanized transgenic mice coexpressing the tetracycline-controlled transactivator and human CYP1B1 gene. *Int J Toxicol*, 24, 157-164.**
86. **Ladner, S.K., Otto, M.J., Barker, C.S., Zaifert, K., Wang, G.H., Guo, J.T., Seeger, C. and King, R.W. (1997) Inducible expression of human hepatitis B virus (HBV) in stably transfected hepatoblastoma cells: a novel system for screening potential inhibitors of HBV replication. *Antimicrob Agents Chemother*, 41, 1715-1720.**
87. **Duerr, J., Gruner, M., Schubert, S.C., Haberkorn, U., Bujard, H. and Mall, M.A. (2010) Use of a new-generation reverse tetracycline transactivator system for quantitative control of conditional gene expression in the murine lung. *Am J Respir Cell Mol Biol*, 44, 244-254.**
88. **Wells, K.D., Foster, J.A., Moore, K., Pursel, V.G. and Wall, R.J. (1999) Codon optimization, genetic insulation, and an rtTA reporter improve performance of the tetracycline switch. *Transgenic Res*, 8, 371-381.**
89. **Urlinger, S., Baron, U., Thellmann, M., Hasan, M.T., Bujard, H. and Hillen, W. (2000) Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc Natl Acad Sci U S A*, 97, 7963-7968.**
90. **Valencik, M.L. and McDonald, J.A. (2001) Codon optimization markedly improves doxycycline regulated gene expression in the mouse heart. *Transgenic Res*, 10, 269-275.**
91. **Park, H.J. and RajBhandary, U.L. (1998) Tetracycline-regulated suppression of amber codons in mammalian cells. *Mol Cell Biol*, 18, 4418-4425.**
92. **Yoshida, Y. and Hamada, H. (1997) Adenovirus-mediated inducible gene expression through tetracycline-controllable transactivator with nuclear localization signal. *Biochem Biophys Res Commun*, 230, 426-430.**
93. **Gould, D.J. and Chernajovsky, Y. (2004) Endogenous GATA factors bind the core sequence of the tetO and influence gene regulation with the tetracycline system. *Mol Ther*, 10, 127-138.**
94. **Williams, T. and Tjian, R. (1991) Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. *Genes Dev*, 5, 670-682.**
95. **Ou, H.L., Huang, Y., Qu, L.J., Xu, M., Yan, J.B., Ren, Z.R., Huang, S.Z. and Zeng, Y.T. (2009) A phiC31 integrase-mediated integration hotspot in favor of transgene expression exists in the bovine genome. *FEBS J*, 276, 155-163.**

96. **Beard, C., Hochedlinger, K., Plath, K., Wutz, A. and Jaenisch, R. (2006) Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis*, 44, 23-28.**

97. **Soriano, P. (1999) Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet*, 21, 70-71.**

98. **Cebula, M., Ochel, A., Hillebrand, U., Pils, M.C., Schirmbeck, R., Hauser, H. and Wirth, D. (2013) An inducible transgenic mouse model for immune mediated hepatitis showing clearance of antigen expressing hepatocytes by CD8+ T cells. *PLoS One*, 8, e68720.**

99. **Sandhu, U., Cebula, M., Behme, S., Riemer, P., Wodarczyk, C., Metzger, D., Reimann, J., Schirmbeck, R., Hauser, H. and Wirth, D. (2010) Strict control of transgene expression in a mouse model for sensitive biological applications based on RMCE compatible ES cells. *Nucleic Acids Res*, 39, e1.**

100. **Kohlhepp, R.L., Hegge, L.F. and Moser, A.R. (2001) The ROSA26 LacZ-neo(R) insertion confers resistance to mammary tumors in Apc(Min/+) mice. *Mamm Genome*, 12, 606-611.**

101. **Kobayashi, T., Kato-Itoh, M., Yamaguchi, T., Tamura, C., Sanbo, M., Hirabayashi, M. and Nakauchi, H. (2012) Identification of rat Rosa26 locus enables generation of knock-in rat lines ubiquitously expressing tdTomato. *Stem Cells Dev*, 21, 2981-2986.**

102. **Irion, S., Luche, H., Gadue, P., Fehling, H.J., Kennedy, M. and Keller, G. (2007) Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nat Biotechnol*, 25, 1477-1482.**

103. **Hermann, M., Maeder, M.L., Rector, K., Ruiz, J., Becher, B., Burki, K., Khayter, C., Aguzzi, A., Joung, J.K., Buch, T. *et al.*(2012) Evaluation of OPEN zinc finger nucleases for direct gene targeting of the ROSA26 locus in mouse embryos. *PLoS One*, 7, e41796.**

104. **Perez-Pinera, P., Ousterout, D.G., Brown, M.T. and Gersbach, C.A. (2011) Gene targeting to the ROSA26 locus directed by engineered zinc finger nucleases. *Nucleic Acids Res*, 40, 3741-3752.**

105. **Kruse, N., Spencer, S., Wirth, D. . (in press (2014)) In Hauser, H., Wagner, R. (ed.), *Animal Cell Biotechnology in Biologics Production*. DeGruyter, Berlin.**

106. **Turan, S. and Bode, J. (2011) Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications. *FASEB J*, 25, 4088-4107.**

107. **Qiao, J., Oumard, A., Wegloehner, W. and Bode, J. (2009) Novel tag-and-exchange (RMCE) strategies generate master cell clones with predictable and stable transgene expression properties. *J Mol Biol*, 390, 579-594.**

108. **Wirth, D., Gama-Norton, L., Riemer, P., Sandhu, U., Schucht, R. and Hauser, H. (2007) Road to precision: recombinase-based targeting technologies for genome engineering. *Curr Opin Biotechnol*, 18, 411-419.**

109. **Gama-Norton, L., Botezatu, L., Herrmann, S., Schweizer, M., Alves, P.M., Hauser, H. and Wirth, D. (2011) Lentivirus production is influenced by SV40 large T-antigen and chromosomal integration of the vector in HEK293 cells. *Hum Gene Ther*, 22, 1269-1279.**

110. **Schucht, R., Lydford, S., Andzinski, L., Zauers, J., Cooper, J., Hauser, H., Wirth, D. and May, T. (2011) Rapid establishment of G-protein-coupled receptor-expressing cell lines by site-specific integration. *J Biomol Screen*, 16, 323-331.**

111. **Wilke, S., Groebe, L., Maffenbeier, V., Jager, V., Gossen, M., Josewski, J., Duda, A., Polle, L., Owens, R.J., Wirth, D. *et al.*(2011) Streamlining homogeneous glycoprotein production for biophysical and structural applications by targeted cell line development. *PLoS One*, 6, e27829.**

112. **Wiberg, F.C., Rasmussen, S.K., Frandsen, T.P., Rasmussen, L.K., Tengbjerg, K., Coljee, V.W., Sharon, J., Yang, C.Y., Bregenholt, S., Nielsen, L.S. *et al.* (2006) Production of target-specific recombinant human polyclonal antibodies in mammalian cells. *Biotechnology and bioengineering*, 94, 396-405.**

113. Chen, C.M., Krohn, J., Bhattacharya, S. and Davies, B. (2011) A comparison of exogenous promoter activity at the ROSA26 locus using a PhiiC31 integrase mediated cassette exchange approach in mouse ES cells. *PLoS One*, 6, e23376.

114. Schucht, R., Wirth, D. and May, T. (2009) Precise regulation of transgene expression level and control of cell physiology. *Cell Biol Toxicol*, 26, 29-42.

115. Schucht, R., Coroadinha, A.S., Zanta-Boussif, M.A., Verhoeyen, E., Carrondo, M.J., Hauser, H. and Wirth, D. (2006) A new generation of retroviral producer cells: predictable and stable virus production by Flp-mediated site-specific integration of retroviral vectors. *Mol Ther*, 14, 285-292.

116. Nehlsen, K., da Gama-Norton, L., Schucht, R., Hauser, H. and Wirth, D. (2012) Towards rational engineering of cells: Recombinant gene expression in defined chromosomal loci. *BMC Proc*, 5 Suppl 8, O6.

117. Turan, S., Zehe, C., Kuehle, J., Qiao, J. and Bode, J. (2012) Recombinase-mediated cassette exchange (RMCE) - a rapidly-expanding toolbox for targeted genomic modifications. *Gene*, 515, 1-27.

118. Brooks, A.R., Harkins, R.N., Wang, P., Qian, H.S., Liu, P. and Rubanyi, G.M. (2004) Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle. *J Gene Med*, 6, 395-404.

119. Meilinger, D., Fellinger, K., Bultmann, S., Rothbauer, U., Bonapace, I.M., Klinkert, W.E., Spada, F. and Leonhardt, H. (2009) Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep*, 10, 1259-1264.

120. Mehta, A.K., Majumdar, S.S., Alam, P., Gulati, N. and Brahmachari, V. (2009) Epigenetic regulation of cytomegalovirus major immediate-early promoter activity in transgenic mice. *Gene*, 428, 20-24.

121. Yang, Y., Mariati, Chusainow, J. and Yap, M.G. (2010) DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. *J Biotechnol*, 147, 180-185.

122. Strunnikova, M., Schagdarsurengin, U., Kehlen, A., Garbe, J.C., Stampfer, M.R. and Dammann, R. (2005) Chromatin inactivation precedes de novo DNA methylation during the progressive epigenetic silencing of the RASSF1A promoter. *Mol Cell Biol*, 25, 3923-3933.

123. Mutskov, V. and Felsenfeld, G. (2004) Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J*, 23, 138-149.

124. Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*, 33 Suppl, 245-254.

125. Guillemette, B., Drogaris, P., Lin, H.H., Armstrong, H., Hiragami-Hamada, K., Imhof, A., Bonneil, E., Thibault, P., Verreault, A. and Festenstein, R.J. (2011) H3 lysine 4 is acetylated at active gene promoters and is regulated by H3 lysine 4 methylation. *PLoS Genet*, 7, e1001354.

126. Angel, A., Song, J., Dean, C. and Howard, M. (2011) A Polycomb-based switch underlying quantitative epigenetic memory. *Nature*, 476, 105-108.

127. Mariani, M.R., Carpaneto, E.M., Ulivi, M., Allfrey, V.G. and Boffa, L.C. (2003) Correlation between butyrate-induced histone hyperacetylation turn-over and c-myc expression. *J Steroid Biochem Mol Biol*, 86, 167-171.

128. Boudadi, E., Stower, H., Halsall, J.A., Rutledge, C.E., Leeb, M., Wutz, A., O'Neill, L.P., Nightingale, K.P. and Turner, B.M. (2013) The histone deacetylase inhibitor sodium valproate causes limited transcriptional change in mouse embryonic stem cells but selectively overrides Polycomb-mediated Hoxb silencing. *Epigenetics Chromatin*, 6, 11.

129. **Wulhfard, S., Baldi, L., Hacker, D.L. and Wurm, F. (2010) Valproic acid enhances recombinant mRNA and protein levels in transiently transfected Chinese hamster ovary cells. *J Biotechnol*, 148, 128-132.**

130. **Nehlsen, K., Schucht, R., da Gama-Norton, L., Kromer, W., Baer, A., Cayli, A., Hauser, H. and Wirth, D. (2009) Recombinant protein expression by targeting pre-selected chromosomal loci. *BMC Biotechnol*, 9, 100.**

131. **Coroadinha, A.S., Schucht, R., Gama-Norton, L., Wirth, D., Hauser, H. and Carrondo, M.J. (2006) The use of recombinase mediated cassette exchange in retroviral vector producer cell lines: predictability and efficiency by transgene exchange. *J Biotechnol*, 124, 457-468.**

132. **Kunert, R. and Casanova, E. Recent advances in recombinant protein production: (2013) BAC-based expression vectors, the bigger the better. *Bioengineered*, 4, 258-261.**

133. **Weber, L.W., Boll, M. and Stampfl, A. (2003) Hepatotoxicity and mechanism of action of haloalkanes: carbon tetrachloride as a toxicological model. *Crit Rev Toxicol*, 33, 105-136.**

134. **Rebollo, R., Miceli-Royer, K., Zhang, Y., Farivar, S., Gagnier, L. and Mager, D.L. (2012) Epigenetic interplay between mouse endogenous retroviruses and host genes. *Genome Biol*, 13, R89.**

135. **Orend, G., Kuhlmann, I. and Doerfler, W. (1991) Spreading of DNA methylation across integrated foreign (adenovirus type 12) genomes in mammalian cells. *J Virol*, 65, 4301-4308.**

136. **Turker, M.S. (2002) Gene silencing in mammalian cells and the spread of DNA methylation. *Oncogene*, 21, 5388-5393.**

137. **Masui, S., Shimosato, D., Toyooka, Y., Yagi, R., Takahashi, K. and Niwa, H. (2005) An efficient system to establish multiple embryonic stem cell lines carrying an inducible expression unit. *Nucleic Acids Res*, 33, e43.**

138. **Backman, C.M., Zhang, Y., Malik, N., Shan, L., Hoffer, B.J., Westphal, H. and Tomac, A.C. (2009) Generalized tetracycline induced Cre recombinase expression through the ROSA26 locus of recombinant mice. *J Neurosci Methods*, 176, 16-23.**

139. **Wortge, S., Eshkind, L., Cabezas-Wallscheid, N., Lakaye, B., Kim, J., Heck, R., Abassi, Y., Diken, M., Sprengel, R. and Bockamp, E. (2010) Tetracycline-controlled transgene activation using the ROSA26-iM2-GFP knock-in mouse strain permits GFP monitoring of DOX-regulated transgene-expression. *BMC Dev Biol*, 10, 95.**

140. **Jahner, D. and Jaenisch, R. (1985) Retrovirus-induced de novo methylation of flanking host sequences correlates with gene inactivity. *Nature*, 315, 594-597.**

141. **Dar, R.D., Razooky, B.S., Singh, A., Trimeloni, T.V., McCollum, J.M., Cox, C.D., Simpson, M.L. and Weinberger, L.S. (2012) Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc Natl Acad Sci U S A*, 109, 17454-17459.**

142. **Coulon, A., Chow, C.C., Singer, R.H. and Larson, D.R. (2013) Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat Rev Genet*, 14, 572-584.**

143. **Skupsky, R., Burnett, J.C., Foley, J.E., Schaffer, D.V. and Arkin, A.P. (2010) HIV promoter integration site primarily modulates transcriptional burst size rather than frequency. *PLoS Comput Biol*, 6.**

144. **Elgin, S.C. and Reuter, G. (2013) Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harb Perspect Biol*, 5, a017780.**

145. **Jeffery, D.C., Wyse, B.A., Rehman, M.A., Brown, G.W., You, Z., Oshidari, R., Masai, H. and Yankulov, K.Y. (2013) Analysis of epigenetic stability and conversions in Saccharomyces cerevisiae reveals a novel role of CAF-I in position-effect variegation. *Nucleic Acids Res*, 41, 8475-8488.**

146. **Yankulov, K. Dynamics and stability: (2013) epigenetic conversions in position effect variegation. *Biochem Cell Biol*, 91, 6-13.**

147. **Schotta, G., Ebert, A., Dorn, R. and Reuter, G. (2003) Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila. *Semin Cell Dev Biol*, 14, 67-75.**

148. **Hsieh, J. and Fire, A. (2000) Recognition and silencing of repeated DNA. *Annu Rev Genet*, 34, 187-204.**
149. **Wong, E.T., Kolman, J.L., Li, Y.C., Mesner, L.D., Hillen, W., Berens, C. and Wahl, G.M. (2005) Reproducible doxycycline-inducible transgene expression at specific loci generated by Cre-recombinase mediated cassette exchange. *Nucleic Acids Res*, 33, e147.**
150. **Giel-Moloney, M., Krause, D.S., Chen, G., Van Etten, R.A. and Leiter, A.B. (2007) Ubiquitous and uniform in vivo fluorescence in ROSA26-EGFP BAC transgenic mice. *Genesis*, 45, 83-89.**
151. **Sun, J., Shao, Z., Zhao, H., Nair, N., Wen, F. and Xu, J.H. (2012) Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in Saccharomyces cerevisiae. *Biotechnol Bioeng*, 109, 2082-2092.**
152. **Haenebalcke, L., Goossens, S., Naessens, M., Kruse, N., Ghahremani, M.F., Bartunkova, S., Haigh, K., Pieters, T., Dierickx, P., Drogat, B. *et al.*(2013) Efficient ROSA26-Based Conditional and/or Inducible Transgenesis Using RMCE-Compatible F1 Hybrid Mouse Embryonic Stem Cells. *Stem Cell Rev*.**
153. **Zheng, J.M., He, Y. and Fu, J.L. (2001) [The study of a HPRT locus-specific transgenic method based on FLP recombinase mediated cassette exchange]. *Shi Yan Sheng Wu Xue Bao*, 34, 279-282.**
154. **Haenebalcke, L., Goossens, S., Dierickx, P., Bartunkova, S., D'Hont, J., Haigh, K., Hochepied, T., Wirth, D., Nagy, A. and Haigh, J.J. (2013) The ROSA26-iPSC mouse: a conditional, inducible, and exchangeable resource for studying cellular (De)differentiation. *Cell Rep*, 3, 335-341.**
155. **Strathdee, D., Ibbotson, H. and Grant, S.G. (2006) Expression of transgenes targeted to the Gt(ROSA)26Sor locus is orientation dependent. *PLoS One*, 1, e4.**
156. **Cedar, H. and Bergman, Y. (2009) Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet*, 10, 295-304.**
157. **Stancheva, I. (2005) Caught in conspiracy: cooperation between DNA methylation and histone H3K9 methylation in the establishment and maintenance of heterochromatin. *Biochem Cell Biol*, 83, 385-395.**
158. **McBurney, M.W., Mai, T., Yang, X. and Jardine, K. (2002) Evidence for repeat-induced gene silencing in cultured Mammalian cells: inactivation of tandem repeats of transfected genes. *Exp Cell Res*, 274, 1-8.**
159. **Eszterhas, S.K., Bouhassira, E.E., Martin, D.I. and Fiering, S. (2002) Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. *Mol Cell Biol*, 22, 469-479.**
160. **Ellis, J. (2005) Silencing and variegation of gammaretrovirus and lentivirus vectors. *Hum Gene Ther*, 16, 1241-1246.**
161. **Weiner, A., Chen, H.V., Liu, C.L., Rahat, A., Klien, A., Soares, L., Gudipati, M., Pfeffner, J., Regev, A., Buratowski, S. *et al.*(2012) Systematic dissection of roles for chromatin regulators in a yeast stress response. *PLoS Biol*, 10, e1001369.**
162. **Smith, K.T. and Workman, J.L. (2012) Chromatin proteins: key responders to stress. *PLoS Biol*, 10, e1001371.**
163. **Garcia-Ramirez, M., Rocchini, C. and Ausio, J. (1995) Modulation of chromatin folding by histone acetylation. *J Biol Chem*, 270, 17923-17928.**
164. **Zhu, P., Aller, M.I., Baron, U., Cambridge, S., Bausen, M., Herb, J., Sawinski, J., Cetin, A., Osten, P., Nelson, M.L. *et al.* (2007) Silencing and un-silencing of tetracycline-controlled genes in neurons. *PLoS One*, 2, e533.**
165. **Manuelidis, L. (1991) Heterochromatic features of an 11-megabase transgene in brain cells. *Proc Natl Acad Sci U S A*, 88, 1049-1053.**
166. **Chevalier-Mariette, C., Henry, I., Montfort, L., Capgras, S., Forlani, S., Muschler, J. and Nicolas, J.F. (2003) CpG content affects gene silencing in mice: evidence from novel transgenes. *Genome Biol*, 4, R53.**

167. Papadakis, E.D., Nicklin, S.A., Baker, A.H. and White, S.J. (2004) Promoters and control elements: designing expression cassettes for gene therapy. *Curr Gene Ther*, 4, 89-113.

168. Di Ianni, M., Terenzi, A., Perruccio, K., Ciurnelli, R., Lucheroni, F., Benedetti, R., Martelli, M.F. and Tabilio, A. (1999) 5-Azacytidine prevents transgene methylation in vivo. *Gene Ther*, 6, 703-707.

169. Jaubert, J., Patel, S., Cheng, J. and Segre, J.A. (2004) Tetracycline-regulated transactivators driven by the involucrin promoter to achieve epidermal conditional gene expression. *J Invest Dermatol*, 123, 313-318.

170. Liu, X., Alexander, V., Vijayachandra, K., Bhogte, E., Diamond, I. and Glick, A. (2001) Conditional epidermal expression of TGFbeta 1 blocks neonatal lethality but causes a reversible hyperplasia and alopecia. *Proc Natl Acad Sci U S A*, 98, 9139-9144.

171. Hwang, D.Y., Chae, K.R., Shin, D.H., Hwang, J.H., Lim, C.H., Kim, Y.J., Kim, B.J., Goo, J.S., Shin, Y.Y., Jang, I.S. *et al.* (2001) Xenobiotic response in humanized double transgenic mice expressing tetracycline-controlled transactivator and human CYP1B1. *Arch Biochem Biophys*, 395, 32-40.

172. McCreath, K.J., Howcroft, J., Campbell, K.H., Colman, A., Schnieke, A.E. and Kind, A.J. (2000) Production of gene-targeted sheep by nuclear transfer from cultured somatic cells. *Nature*, 405, 1066-1069.

173. Wan, M., Gu, H., Wang, J., Huang, H., Zhao, J., Kaundal, R.K., Yu, M., Kushwaha, R., Chaiyachati, B.H., Deerhake, E. *et al.*(2013) Inducible mouse models illuminate parameters influencing epigenetic inheritance. *Development*, 140, 843-852.

174. Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V. *et al.*(2012) In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet*, 8, e1002750.

175. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.

176. Papapetrou, E.P., Lee, G., Malani, N., Setty, M., Riviere, I., Tirunagari, L.M., Kadota, K., Roth, S.L., Giardina, P., Viale, A. *et al.* (2011) Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells. *Nature biotechnology*, 29, 73-78.

177. Sadelain, M., Papapetrou, E.P. and Bushman, F.D. (2012) Safe harbours for the integration of new DNA in the human genome. *Nature reviews. Cancer*, 12, 51-58.

178. Nunez, E., Fu, X.D. and Rosenfeld, M.G. (2009) Nuclear organization in the 3D space of the nucleus - cause or consequence? *Current opinion in genetics & development*, 19, 424-436.

179. Sadelain, M., Papapetrou, E.P. and Bushman, F.D. (2011) Safe harbours for the integration of new DNA in the human genome. *Nat Rev Cancer*, 12, 51-58.

9APPENDIX

9.1 LIST OF ABBREVATIONS

Abbreviations	Explanations
µ	mikro; 10-6
Ampr	Ampicillin resistance gene
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
Aza	azacytidine
BCA	bicinchoninic acid
BiTet	bidirectional Tet promoter
BLI	bioluminescent imaging
bp	base pair
BSA	bovine serum albumin
CCL4	carbon tetrachloride
CMV	cytomegalovirus
CGI	CpG islands
Deci	decitabine
DMEM	Dulbecco's modified Eagle's medium
Dnmt	DNA methyltransferase
Dnmti	DNA methyltransferase inhibitor
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
Dox	doxycycline

DTT	dithiothreitol
eGFP	enhanced green fluorescent protein
eGFP/Neo	fusion protein of eGFP und neomycine resistance gene
et al.	et alii
FACS	fluorescence-activated cell sorting
FCS	fetal calf serum
FSC	forward scatter
g	gram
G418	aminoglycoside-2'-deoxystreptine (gentamycin derivative)
GFP	green fluorescent protein
GOI	gene of interest
h	hour(s)
HDAC	Histone deacetylase
HDACi	Histone deacetylase inhibitors
i.p.	intraperitoneal
IRES	internal ribosomal entry site
k	kilo; 103
kb	1000 base(s)
KO	knock out
kV	1000 volt(s)
l	liter(s)
LTR	long terminal repeat
mES	murine embryonic stem cel
min	minute(s)
ml	milliliter(s)

mRNA	messenger RNA
NaB	sodium butyrate
Neo	neomycinephosphotransferase
ori	origin of replication
p	piko; 10-9
p53	tumor suppressor protein 53
pA	polyadenylation signal
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PE	R-phycoerythrin
PGK	phosphoglycerate kinase
RFP	red fluorescent protein
RMCE	recombinase mediated cassette exchange
RNA	ribonucleic acid
RNase	ribonuclease
rpm	rounds per minute
rre	Rev responsive element
RT	room temperature
rtTA	reverse transactivator of the tet system
s	second(s)
SDS	sodium dodecyl sulfate
SIN	self inactivating
SMC	smooth muscle cell
SSC	sideward scatter
SV40	Simian Virus 40

Tab.	Table
TAg	SV40 Large Tumor Antigen
TE	Trypsin- EDTA
Tet	Tetracycline
tetO	operator sequence of tetracyclin resistance system
Tet-Off	tet dependent expression system
Tet-On	tet dependent expression system
TetR	tetracycline Repressor domain
Tris	trishydroxymethylaminomethane
tTA	transactivator of the tet system
U	unit
VPA	valproic acid
v/v	volume/volume; percent by volume
w.t.	wild type
w/v	weight/volume; percent by weight

9.2 LIST OF FIGURES

9.3 LIST OF TABLES